University of Bath

PHD

Stationary phase genes of Saccharomyces cerevisiae

Henstock, Mark Richard

Award date:
2004

Awarding institution:
University of Bath

Link to publication

Alternative formats
If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

Copyright of this thesis rests with the author. Access is subject to the above licence, if given. If no licence is specified above, original content in this thesis is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC-ND 4.0) Licence (https://creativecommons.org/licenses/by-nc-nd/4.0/). Any third-party copyright material present remains the property of its respective owner(s) and is licensed under its existing terms.

Take down policy
If you consider content within Bath's Research Portal to be in breach of UK law, please contact: openaccess@bath.ac.uk with the details. Your claim will be investigated and, where appropriate, the item will be removed from public view as soon as possible.

Download date: 04. May. 2023

Stationary phase genes of

Saccharomyces cerevisiae

submitted by Mark Richard Henstock

for the degree of PhD

of the University of Bath

2004

COPYRIGHT

Attention is drawn to the fact that copyright of this thesis rests with its author.

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without the prior written consent of the author.

This thesis may be made available for consultation within the University Library and may be photocopied or lent to other libraries for the purposes of consultation.

UMI Number: U206982

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

UMI U206982
Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.
All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

55 3 1 MAY 2006
Ph.D.

Acknowledgements

Thanks to Alan Wheals for all his help and advice, and the BBSRC for funding my studentship. Thanks also go to Elizabeth Williams and Araxi Urrutia Odabachian for their invaluable help with the work on IRES. I'd also like to acknowledge all my friends and family who kept me sane and calm throughout my PhD, in particular my fiancée Janine for her patience and support.

Table of Contents

CHAPTER 1

INTRODUCTION

Physiology of *Saccharomyces cerevisiae*

The yeast *S. cerevisiae* is a eukaryotic cell, of about 5–10 μm in diameter (Elert; Martinez). The cell is surrounded by a phospholipid bilayer plasma membrane that in turn is surrounded by the periplasm and a cell wall consisting mainly of glucan and mannan (Martinez). Inside the cell is a single nucleus, the Golgi apparatus, an endoplasmic reticulum, one or more mitochondria, one or more vacuoles and peroxisomes (Walker 1998).

Like other eukaryotes the *S. cerevisiae* cell cycle is split into four conventional phases: G1, S, G2 and M (Pringle and Hartwell 1981; Rupes 2002; Wiebe 2003). Vegetative replication in *S. cerevisiae*, in both haploid and diploid states, occurs by a genetically identical daughter cell budding from a mother cell (Pringle and Hartwell 1981). Yeast cells have an average replicative lifespan, as measured by the number of daughter cells produced by an actively dividing mother cell, of about 20 to 30 generations (Ashrafi, Sinclair et al. 1999; Nestlbacher, Laun et al. 2000; Fröhlich and Madeo 2001). Multiploid *S. cerevisiae* cells can undergo meiosis and ascospore formation. This generally occurs under conditions of nitrogen starvation in the presence of a nonfermentable carbon source. When in the haploid form a *MAT***a** and a *MAT*α mating type cell combine to form a single diploid cell immediately before ascus formation (Martinez; Esposito and Klapholz 1981).

Stationary phase

Stationary phase is the arrest of growth and division by microorganisms in response to starvation conditions. This is an important adaptive response because a continual state of high nutrient availability cannot be guaranteed and by halting growth and division cells can

conserve what energy and material stores they have so that they can survive the period of starvation (Kolter, Siegele et al. 1993; Werner-Washburne, Braun et al. 1993; Werner-Washburne, Braun et al. 1996; Herman 2002).

At its most basic, entry into stationary phase is the switch from a state of growth to non-growth. Closer examination of this process however reveals a much more complicated, highly regulated process. When growing on the rich, glucose-based medium YPD cells of *S. cerevisiae* will preferentially grow by fermentation even in the presence of oxygen (Lagunas 1986; Werner-Washburne, Braun et al. 1993; Walker 1998). During fermentation *S. cerevisiae* transforms one molecule of D-glucose, through various intermediates, into two molecules of pyruvate. The pyruvate formed is then decarboxylated to ethanol and excreted from the cell (Diagram 1-1). There is a net production of two molecules of ATP for the fermentation of one molecule of glucose into two molecules of ethanol (Lagunas 1986; Walker 1998).

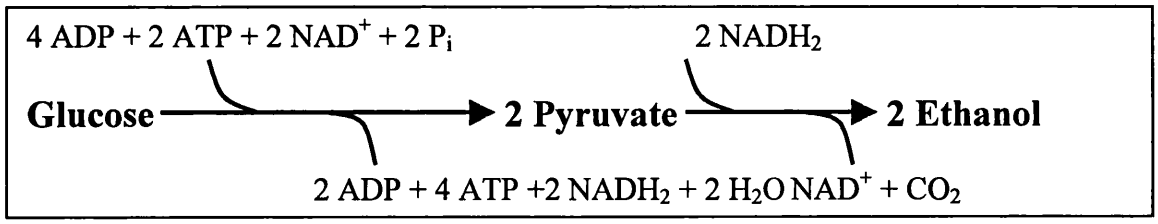


Diagram 1-1: The fermentation of glucose in *S. cerevisiae*.

When glucose becomes exhausted there is a temporary pause in the cell cycle, known as the diauxic shift. During the post-diauxic phase growth is respiratory and uses the products of fermentation as well as other carbon sources (Werner-Washburne, Braun et al. 1993; Werner-Washburne, Braun et al. 1996). Like fermentation, respiration converts one molecule of D-glucose into two molecules of pyruvate and two molecules of ATP. However, instead of decarboxylation to ethanol, the pyruvate is oxidized to carbon dioxide

and water using oxygen as the terminal electron accepter (Diagram 1-2). The ATP yield of respiration is 24 to 34 molecules of ATP produced per molecule of D-glucose respired (Lagunas 1986).

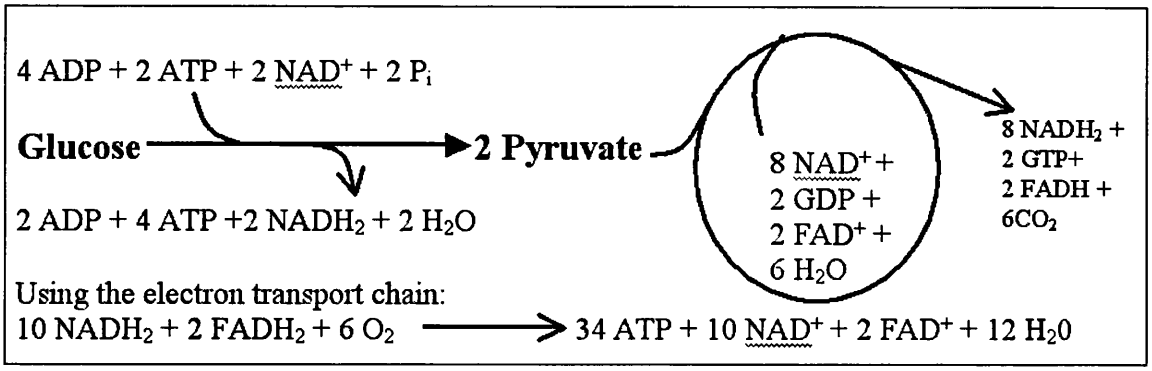


Diagram 1-2: The respiration of glucose in *S. cerevisiae*.

When the carbon sources fuelling respiratory growth are exhausted cellular growth and division are halted and cells enter stationary phase (Diagram 1-3) (Lagunas 1986; Werner-Washburne, Braun et al. 1993; Fuge, Braun et al. 1994; Werner-Washburne, Braun et al. 1996). During stationary phase glycogen and trehalose are respired to provide energy (Lillie and Pringle 1980; Longo, Gralla et al. 1996; Roy and Ghosh 1998; Silljé, Paalman et al. 1999; Samokhvalov, Ignatov et al. 2004). During respiration both glycogen and trehalose are broken down and channelled into the glycolytic pathway (glucose to pyruvate). Glycogen is first broken down into glucose-1-phosphate, which is converted into glucose-6-phosphate, and trehalose is converted into directly into D-Glucose (Diagram 1-4).

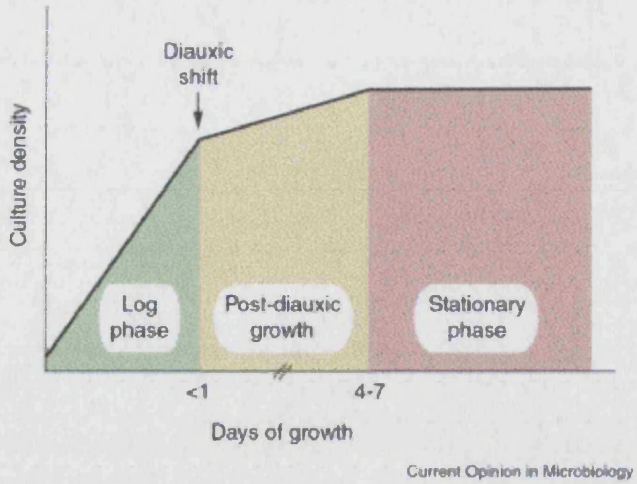


Diagram 1-3: line chart showing growth from exponential growth to stationary phase (figure 1 from (Herman 2002).

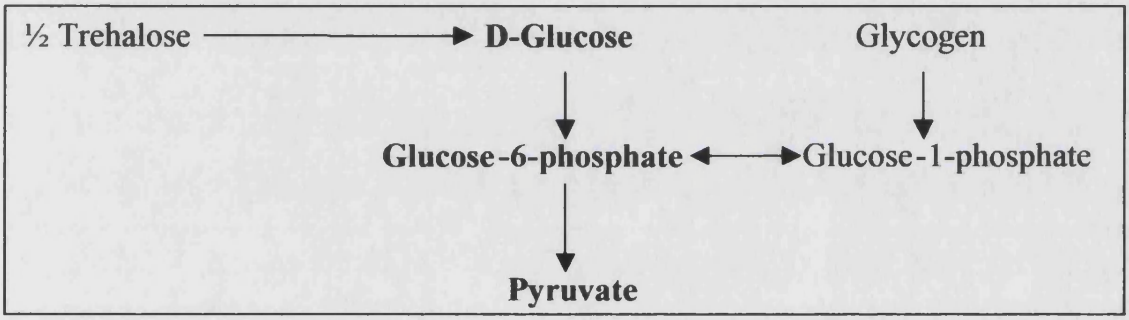


Diagram 1-4: Respiration of trehalose and glycogen during stationary phase.

Stationary Phase in the cell cycle

When arrested in stationary phase *S. cerevisiae* cells are unbudded and have an unreplicated DNA content characteristic of cells in the G_1 phase of the cell cycle (Pringle and Hartwell 1981; Werner-Washburne, Braun et al. 1993; Herman 2002). This suggests that when arrested in stationary phase cells are halted during the G_1 stage of the cell cycle. The cold sensitive *gcs1 sed1* mutant is unable to re-enter the cell cycle from stationary phase at the restrictive temperature of 14°C. When stationary phase mutants at the restrictive temperature are diluted into fresh medium they lose stationary phase characteristics and obtain characteristics of actively dividing cells, but are unable to divide.

Using response to mating pheromone and morphological features the re-entry block caused by this mutation was found to occur during G_1, before the cell cycle step 'start' (Drebot, Johnston et al. 1987). Unlike conditional mutants that halt at 'start' and then released, cells exiting from stationary phase are not synchronised (Werner-Washburne, Wylie et al. 2002). 'Start' is a point in the G_1 phase at which cell proliferation is regulated (Pringle and Hartwell 1981). Various conditions must be met to proceed past this point, undergo DNA synthesis and produce a bud. These include adequate nutrients in the medium and attainment of a critical size (Sudbery, Goodey et al. 1980).

Work with conditional mutants has shown that stationary phase is not part of G_1, but is a separate distinct state. Cells that enter and exit stationary phase take longer to traverse G_1 than those cells that do not enter stationary phase. Suggesting that entering and exiting stationary phase require the completion of a process that cycling cells are not required to undertake (Diagram 1-5A) (Johnston, Pringle et al. 1977). Arrest in the G_1 phase of the cell cycle does not appear to be an essential condition of stationary phase. Using conditional mutants to halt cells in different phases of the cell cycle it has been shown that the mechanisms used to enter stationary phase are able to operate in any stage of the cell cycle, not just during G_1 (Wei, Nurse et al. 1993).

Observations of stationary phase cells show that a similar state of arrest is caused by nitrogen, sulphur, and carbon starvation (Drebot, Barnes et al. 1990). However long term maintenance of viability is only elicited by carbon starvation (Granot and Snyder 1991; Granot and Snyder 1993). This has led to a two-state model of growth arrest and stationary phase in response to starvation (Werner-Washburne, Braun et al. 1996). Starvation for any nutrient causes cells to exit from the cell cycle and enter G_0. However only the absence of a carbon source will allow the acquisition of a state where the cell can survive for a prolonged period of time without added nutrients (Diagram 1-5B).

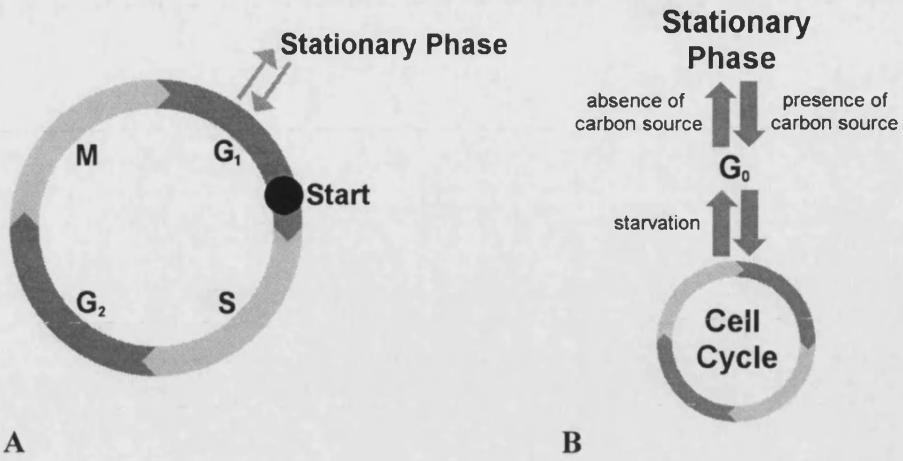


Diagram 1-5: (A) Entry to and exit from stationary phase occurs before the cell cycle point 'start' (Diagram based on Figure 2 from (Werner-Washburne, Braun et al. 1993). (B) Two-state model, allowing for the increased maintenance of viability in stationary phase only seen with starvation for a carbon source (Diagram based on Figure 4 from (Werner-Washburne, Braun et al. 1996).

Characteristics of cells in stationary phase

Certain characteristics are associated with cells that have entered stationary phase due to the gradual depletion of the carbon source. An abrupt removal of the carbon source from the medium does not allow the proper establishment of stationary phase (Martínez-Pastor and Estruch 1996; Jona, Choder et al. 2000). Cells starved for nutrients other than carbon, such as nitrogen or sulphur, show all the characteristics of a carbon starved cell except the ability to survive for prolonged periods in starvation conditions (Drebot, Barnes et al. 1990; Granot and Snyder 1991; Granot and Snyder 1993).

Cells entering stationary phase undergo several visual changes. They appear refractile (phase bright) under phase contrast microscopy and are unbudded. The cell walls become thicker and less porous and are also more resistant to digestion by the enzymes zymolyase and glusulase. Under some conditions cells have more rounded and numerous

mitochondria (Stevens 1981; Bugeja, Piggott et al. 1982; de Nobel, Klis et al. 1990; Werner-Washburne, Braun et al. 1993; Herman 2002).

There are also internal cellular changes that have been linked to entry into stationary phase. Lipid vesicles become more numerous, proteases accumulate in certain subcellular locations and electron dense material accumulates in the vacuole. The DNA content of a stationary phase cell is unreplicated and the chromosomes become folded in a characteristic manner (Piñon 1978; Achstetter, Ehmann et al. 1983; Jones 1984; Werner-Washburne, Braun et al. 1993).

Triacylglycerol synthesis increases, while total phospholipid synthesis decreases and the relative concentrations of phosphotidylinositiol and phosphotidylserine change. Mutants of Opi3p lose viability when in exhausted medium. This protein functions in the final methylation reaction during phophatidylcholine synthesis which suggests that the membrane plays an important part in stationary phase (McGraw and Henry 1989; Werner-Washburne, Braun et al. 1993).

Glycogen and trehalose (Diagram 1-6) accumulate in the cytoplasm and are slowly metabolised during stationary phase (Werner-Washburne, Braun et al. 1993; Herman 2002; Jørgensen, Olsson et al. 2002). The pattern of accumulation and utilization of glycogen and trehalose are different, suggesting that the two carbohydrates fulfil different roles. Glycogen starts to accumulate before the diauxic shift while trehalose doesn't start to accumulate until the diauxic shift, and the levels of both drop as stationary phase proceeds. The patterns suggest that glycogen is being used as an energy source during respiratory adaptation and stationary phase. While trehalose is used as an energy source during stationary phase only and is probably primarily used during stationary phase as a stress protectant (Lillie and Pringle 1980; Werner-Washburne, Braun et al. 1993; Roy and Ghosh 1998).

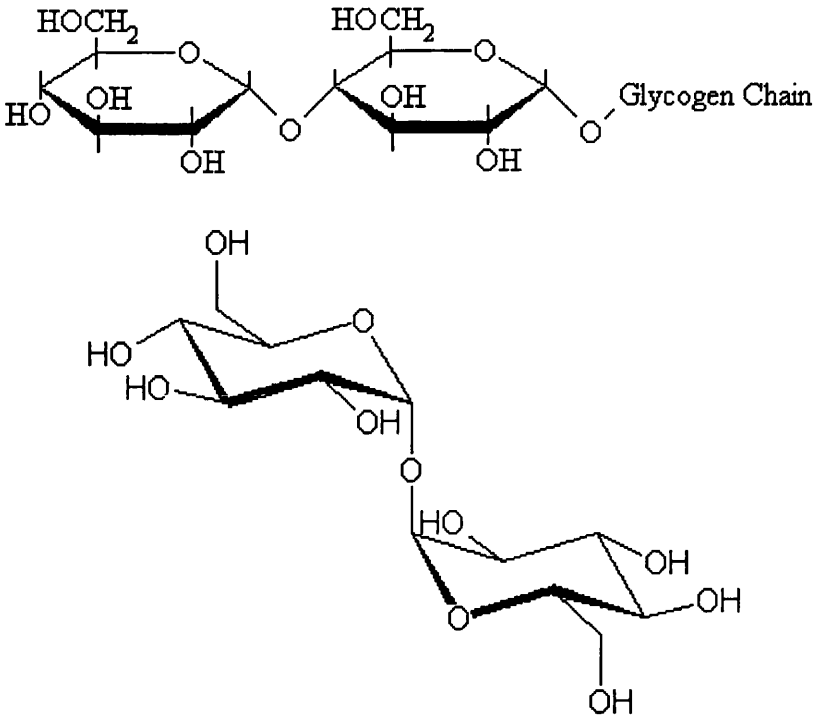


Diagram 1-6: Chemical structures of glycogen (top) and trehalose (bottom). Chemical structures obtained from http://www.gwu.edu/~mpb/polysacc.htm and www.med.unibs.it/ ~marchesi/trehalose.html respectively.

The concentrations in the stationary phase cell of glycogen and trehalose have been observed to affect cell viability. If a cell cannot accumulate glycogen and trehalose as it adapts to starvation conditions stationary phase the viability of the cell during stationary phase is reduced (Lillie and Pringle 1980; Slaughter and Nomura 1992; Silljé, Paalman et al. 1999). It has also been observed that during stationary phase the exhaustion of cellular stores of glycogen and trehalose is associated with a loss in cellular viability (Samokhvalov, Ignatov et al. 2004).

Removal of *GPH1* (encoding glycogen phosphorylase) or over-expression of *GSY2* (encoding glycogen synthase) increases the level of glycogen in a stationary phase cell enhances the ability of cells to survive stationary phase (Pérez-Torrado, Gimeno-Alcaniz et al. 2002). Both *GPH1* and *GSY1* have STRE (Stress Response Element)-like elements in

their promoters that cause them to be regulated by nutrient starvation-induced signalling pathways (Ruis and Schüller 1995).

Cells in stationary phase have increased stress resistance (such as resistance to heat shock) (Werner-Washburne, Braun et al. 1993; Herman 2002). Mutants of Hsp104p (a heat shock protein of the *ClpA/ClpB* family from *Escherichia coli*) suffer increased mortality on heat shock and slight loss of viability in stationary phase (Werner-Washburne, Braun et al. 1993). An over-activation of the heat shock response can, in some cases, increase the chronological lifespan (Harris, MacLean et al. 2001). Mutants of *RVS161* are sensitive to carbon, nitrogen and sulphur starvation and high salt conditions and are unable to grow on nonfermentable carbon sources (Werner-Washburne, Braun et al. 1993). When the mitochondrial superoxide dismutase gene (MnSOD) *SOD2* is deleted, yeast cells display a reduced stationary phase survival phenotype suggesting that the generation of reactive oxygen species (ROS) is detrimental to stationary phase survival (Longo, Liou et al. 1999). At the diauxic shift and during stationary phase several genes encoding antioxidants are induced and cells become more resistant to hydroperoxides indicating that the antioxidant defences of *S. cerevisiae* are under carbon catabolite control (Maris, Assumpção et al. 2001; Cyrne, Martins et al. 2003).

Transcription and translation occur at a reduced rate during stationary phase (Werner-Washburne, Braun et al. 1993), however if this is taken into account over 1000 genes are effectively upregulated (even though the actual rate of mRNA production has fallen) (van de Peppel, Kemmeren et al. 2003). The *cdc33-1* mutation of translation initiation factor eIF-4F causes arrest in stationary phase. This gene may affect stationary phase through the inhibition of translation of certain key genes (Werner-Washburne, Braun et al. 1993). It has also been shown that proteins of the SRB complex (from RNA polymerase II) are required for entry into stationary phase (Herman 2002). Translation of rDNA genes is also affected. Not only is transcription by RNA polymerase I down-regulated but histones H3 and H4 are acetylated by Rpd3p, a histone deacetylase. This

acetylation causes the DNA to shift from an open to a closed (transcription inhibitory) state. In a Δ*rpd3* mutant the transcription of rDNA genes is still reduced in stationary phase though this is due to a reduced number of RNA polymerases (Sandmeier, French et al. 2002).

The *ils1-1* (temperature sensitive) mutation of the isoleucine-tRNA synthetase gene blocks cell proliferation and causes the acquisition of some stationary phase characteristics. This is believed to be a general control problem because *ils1-1 gcn1/gcn2/gcn3* double mutants are not viable (Werner-Washburne, Braun et al. 1993). *Δgcs1* mutants are known to be defective for re-entry to the cell cycle from stationary phase at a restrictive temperature (Drebot, Johnston et al. 1987). Gcs1p is a 39 KDa protein with a 'Zn-finger' motif. Mutation of this motif produces a phenotype identical to the deletion of the whole gene, indicating that the motif is essential for Gcs1p function (Ireland, Johnston et al. 1994). Rpi1p is a transcriptional activator that is localised in the nucleus. The function of this protein seems to be to alter the transcription of several cell wall metabolism proteins in order to fortify the cell wall in preparation for stationary phase (Sobering, Jung et al. 2002).

Ribosome localisation may also have an effect on the correct entry into stationary phase. Work on *Podospora anserina* has shown a transient accumulation of ribosomal proteins and/or subunits in the nucleolus at the beginning of stationary phase (Lalucque and Silar 2000). Rix7p is a novel member of the AAA ATPase superfamily, which associates with 60S ribosomal precursor particles. During exponential growth Rix7p localises throughout the nucleus, but accumulates in the nucleolus during stationary phase (Gadal, Strauß et al. 2001). It is suggested that this is an additional level of regulation involved in ensuring correct cell cycle arrest.

The genes *YPT1* and *PMR1* both are involved in protein transport between the ER and the Golgi body. Ypt1p is a small GTP-binding protein associated with the Golgi body. It is required for transport from the ER through the Golgi body, and is essential for

viability on nitrogen, but not carbon, starvation (Werner-Washburne, Braun et al. 1993). *PMR1* codes for a protein that resembles a calcium ion pump. Loss of this gene compromises viability in stationary phase, and also affects ER to Golgi transition (Werner-Washburne, Braun et al. 1993).

Enzymes involved in protein modification have also been implicated in stationary phase. *ARD1* and *NAT1* mutants affect protein N-terminal acetylation. Both these mutants cease proliferation, but do not acquire stationary phase characteristics when starved (Werner-Washburne, Braun et al. 1993). *UBI4* is the only ubiquitin gene of four in the genome to be induced at entry to and during stationary phase (Finley, Özkaynak et al. 1987; Fraser, Luu et al. 1991; Werner-Washburne, Braun et al. 1993). *UBI4* (encoding polyubiquitin, a heat shock protein) mutants are sensitive to chronic heat stress at 38.5°C but resistant to acute heat stress. *UBI4* mutants seem to be able to enter stationary phase upon starvation but are unable to maintain viability (Finley, Özkaynak et al. 1987). Ubc1p is a ubiquitin-conjugating enzyme and mutants of this gene are slow to resume growth on exiting stationary phase implicating ubiquitination in stationary phase (Werner-Washburne, Braun et al. 1993).

Signal pathways known to be involved with stationary phase in *S. cerevisiae*

Three signalling pathways have been shown to be involved in stationary phase: RAS/cAMP, TOR and *SNF1* protein kinase pathways. The RAS/cAMP and TOR pathways are both sufficient but not essential for entry into stationary phase. A *S. cerevisiae* cell will enter a stationary phase-like state if either of the pathways are disrupted. However if both pathways are disrupted, the cell cannot enter stationary phase (Broach 1991; Schmelzle and Hall 2000; Herman 2002).

The target of rapamycin (TOR) pathway has been implicated as a stationary phase signalling pathway. Cells that are TOR depleted or are treated with rapamycin acquire several physiological characteristics of stationary phase cells (such as down regulation of

transcription and translation, and glycogen accumulation) (Zaragoza, Ghavidel et al. 1998; Hardwick, Kuruvilla et al. 1999; Schmelzle and Hall 2000). The TOR proteins (Tor1p and Tor2p) are activated by the presence of nutrients in the medium, and are inactive under unfavourable growth conditions. The signal pathway components that are upstream of TOR in yeast have yet to be identified.

Downstream of the TOR proteins some of the components of the signal pathway have been identified (Diagram 1-7). When nutrient levels are not low Tor2p (but not Tor1p) activates the Rho1 GTPase. Active Rho1p then starts, via protein kinase C (Pkc1p), a MAP kinase signal cascade which controls cell cycle dependent polarisation of the actin cytoskeleton (Krause and Gray 2002).

When nutrient levels fall, TOR becomes inactive resulting in several different effects. There are changes in the permeases used to import nutrients into the cell, an example being the amino acid transporters Tat2p and Gap1p. In response to nutrient deprivation the high affinity tryptophan transporter Tat2p is degraded and the general amino acid permease Gap1p is upregulated (Diagram 1-7). The nitrogen and stress transcription factors Gln3p (which is also regulated in response to glucose by the Snf1p signal pathway (Bertram, Choi et al. 2002)), Msn2p and Msn4p become relocated from the cytoplasm into the nucleus where they upregulate genes required for stress conditions. It has been shown that they can bind to upstream activation sequence STRE (core consensus sequence: AGGGG or CCCCT) (Ruis and Schüller 1995). Transcription, using all three RNA polymerases, and translation are also down regulated in response to the inactivation of the TOR pathway.

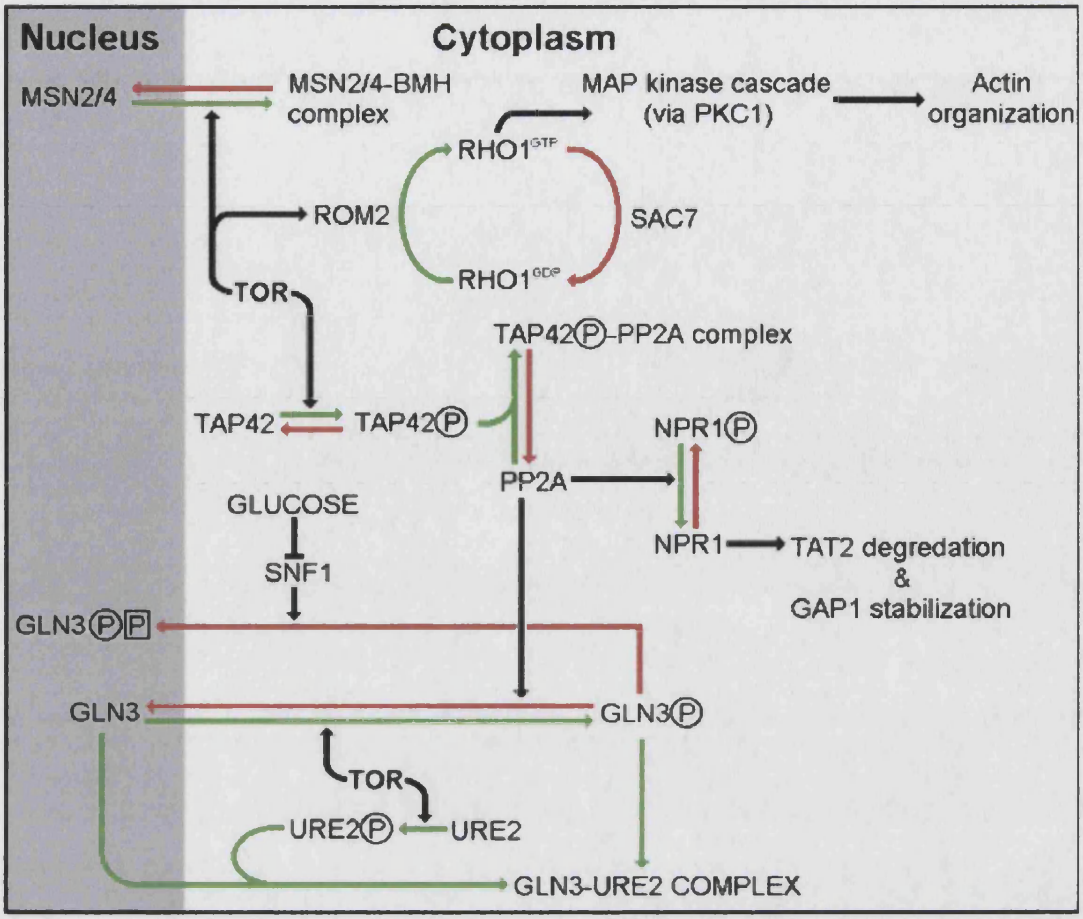


Diagram 1-7: Some of the downstream targets of TOR signalling. Green arrows indicate processes that are favoured when nutrients are not limiting, red arrows those that are favoured under low nutrient conditions. Black arrows indicate which reactions are enhanced by proteins.

The RAS/cAMP pathway consists of two RAS guanine nucleotide-binding proteins (Ras1p and Ras2p). The binding of the RAS proteins to GTP/GDP is regulated by the *IRA* gene products and Cdc25. When bound to GTP, under conditions of high glucose concentrations, RAS proteins are active and activate adenylyl cyclase (Cyr1p) which synthesises cAMP from ATP. The cAMP produced has two possible fates. It can be hydrolysed by the two cAMP phosphodiesterases Pde1p and Pde2p. The second fate of cAMP in the RAS/cAMP pathway is interacting with the Protein Kinase A holoenzyme. Protein Kinase A (Tpk) exists in its inactive form bound to the Bcy1p protein. cAMP

causes the disassociation of Bcy1p from Protein Kinase A. The active Protein Kinase A is free to phosphorylate its targets (Diagram 1-8) (Broach 1991; Werner-Washburne, Braun et al. 1993; Wittenburg and Reed 1996) one being the nuclear localisation signal of Msn2p (described above in the description of the TOR signalling pathway). Phosphorylation of this protein sequence results in the inhibition of its function (Görner, Durchschlag et al. 2002).

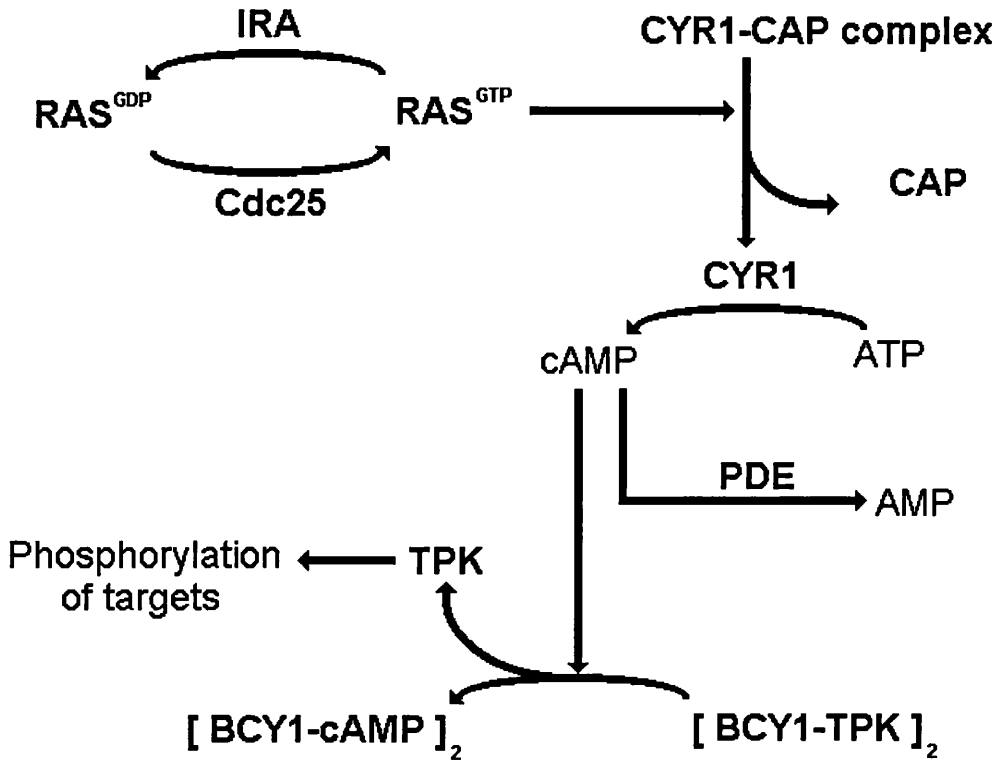


Diagram 1-8: The RAS/cAMP pathway from *S. cerevisiae* (Diagram 1-based on figure 3, from (Werner-Washburne, Braun et al. 1993).

Disruption of the RAS/camp pathway, either by gene knockouts or by gene mutation, affects the cell's ability to enter and survive stationary phase. The effects of these changes can be separated into whether they cause protein kinase A (TPK) either to become activated or to become deactivated (Tatchell 1986; Broach 1991; Werner-

Washburne, Braun et al. 1993). If protein kinase A becomes activated a cell cannot accumulate glycogen or acquire thermotolerance or arrest as unbudded cells and becomes sensitive to starvation (Cannon and Tatchell 1987; Toda, Cameron et al. 1987; Werner-Washburne, Braun et al. 1996). If protein kinase A, becomes deactivated cells hyper-accumulate glycogen, become constitutively thermotolerant and arrest growth as unbudded cells (Kataoka, Broek et al. 1985; Toda, Uno et al. 1985; Werner-Washburne, Braun et al. 1996). Thus loss of the IRA function or overexpression of *CYR1* (Diagram 1-8) will prevent entry into stationary phase by raising cellular cAMP levels (Kataoka, Broek et al. 1985; Tanaka, Matsumoto et al. 1989) whereas loss of the *BCY1* gene will release protein kinase A from cAMP regulation and cause the cell to enter stationary phase (Toda, Uno et al. 1985; Broach 1991).

The Snf1p protein kinase pathway is a third pathway used in the regulation of genes in response to the nutritional status of the cell's medium. One of the major functions of Snf1p is to phosphorylate the transcriptional repressor Mig1p. The level of glucose in the medium determines the status of the Snf1p kinase. Under low glucose conditions Snf1p becomes active and phosphorylates Mig1p, which causes Mig1p to become relocalised from the nucleus to the cytoplasm. This allows the transcription of those genes being repressed by Mig1p (Wilson, Hawley et al. 1996; Carlson 1999).

Mig1p is not the only target of the Snf1p kinase. The localisation of Gln3p is also regulated by phosphorylation by Snf1p (Diagram 1-7) (Bertram, Choi et al. 2002). The activity of Cat8p and Sip4p are also upregulated by Snf1p activity. These two genes are transcriptional activators that upregulate the expression of CSRE (Carbon Source Response Element) -controlled genes, such as gluconeogenic genes (Carlson 1999).

Spa2p is involved in bud site selection but *SPA2* mutants remain more budded than wild type cells when starved and are slightly less thermotolerant (Werner-Washburne, Braun et al. 1993). Slk1p protein kinase mutants are dependent on *SPA2* for growth and fail to respond to nutrient starvation (among other effects). Slk1p and Protein Kinase A

may have antagonising effects in cellular nutrient signalling (Werner-Washburne, Braun et al. 1993). *WHI2* mutants produce daughter cells that are smaller than the mother cell. It has been suggested that Whi2p is part of the nutritional signal transduction pathway and the General Stress Response possibly by acting with Psr1p to dephosphorylate Msn2p (Werner-Washburne, Braun et al. 1993; Kaida, Yashiroda et al. 2002).

Genetics of *S. cerevisiae*

S. cerevisiae can exist in both haploid and diploid forms. Most of the DNA present in the cell is located in the nucleus, but there is also a small genome in the mitochondria. The mitochondrial genome contains some but not all the genes required for mitochondrial function. It is also possible for DNA to exist in *S. cerevisiae* cells in the form of nuclear-located plasmids.

In 1996 *S. cerevisiae* became the first eukaryotic organism to have its genome completely sequenced and it was at the time the largest genome yet sequenced. The sequencing of the genome was undertaken by a consortium of labs from Europe, North America, and Japan (Goffeau, Barrell et al. 1996). The analysis of the sequenced genome showed that it was 13,389 kbp in size, of which 12,068 kbp has been sequenced. A total of 6275 potential open reading frames (ORFs) coding for proteins of more than 99 amino acids in length were identified. However 390 of these potential ORFs were believed to be unlikely to be translated into a protein product (hypothetical ORFs), leaving 5885 potential protein-encoding genes. The genome appears to be very tightly packed, with ORFs comprising almost 70% of the total sequence (Goffeau, Barrell et al. 1996).

The value of 5885 genes in the *S. cerevisiae* genome is not a final number. Work by Ross-Macdonald *et al.* has shown that some genes in the genome may be smaller that the 99 amino acids, the value that was used as the lower limit in the original annotation (Ross-Macdonald, Coelho et al. 1999). There has also been a revision of the annotation of the genome that has reclassified some ORFs as features that are not actually functional

ORFs (Wood, Rutherford et al. 2001). Previous to the start of the genome sequencing project only about 1000 genes had been defined and over half of the annotated ORFs (56%) had never been studied before. Since then almost three-quarters of the ORFs have been assigned an initial functional classification (Goffeau, Barrell et al. 1996; Wood, Rutherford et al. 2001).

The EUROFAN consortium was set up to investigate the function of 1000 unknown genes. The first part of the project (EUROFAN 1) was to create deletions of the 1000 chosen genes, using the *kanMX4* marker in the strain *S. cerevisiae* FY1679 or an isogenic derivative. The creation of a replacement cassette and cognate gene clone and a basic phenotypic analysis of each of the genes was also to be done (Wach, Brachat et al. 1996). Building on EUROFAN 1, the aim of EUROFAN 2 was to conduct a more in-depth analysis of the gene deletions. Each of the labs in the EUROFAN network would conduct a different phenotypic analysis on the set of gene knockouts. The data from each of these labs was then collated, allowing a detailed description of the effects of each gene knockout to be created (http://mips.gsf.de/proj/yeast/CYGD/db/index.html).

Gene knockouts for EUROFAN were created by the method of homologous recombination (Wach, Brachat et al. 1994). A deletion cassette was created for the target gene. The cassette consists of the *kanMX4* marker gene flanked by a pair of sequences homologous to the genomic sequence flanking the target gene. The sequences flanking the *kanMX4* marker were created from a PCR reaction with the *kanMX4* marker and a pair of primers. In the Short-Flanking Homology (SFH) method the primers were a pair specially created 60mer primers (Diagram 1-9). In the Long-Flanking Homology (LFH) method the primers were the result of a pair of PCR reactions using the genomic DNA flanking the target gene and 4 different primers (Diagram 1-9). When *S. cerevisiae* cells are transformed with the deletion cassettes the regions of flanking regions to the *kanMX4* marker allow recombination to occur specifically with the flanking regions of the target

gene. This results in the excision of the target gene and its replacement with the *kanMX4* marker (Wach, Brachat et al. 1996).

Systematically creating gene knockouts provides a useful tool for undertaking a screen of all the genes in the genome for genes that are required for a process. Using a PCR based strategy all known or suspected genes were to be accurately targeted and knocked out. The knockout can be targeted so that either the entire gene is removed or less than the whole gene to avoid affecting an overlapping gene or genomic feature.

A less work-intensive method of screening gene knockouts can be achieved using transposons. A transposon will integrate itself into the genome and will disrupt gene function if it integrates itself into the middle of an ORF. Screening a large number of disruptants for a specific phenotype associated with the disruption of a gene essential for the process of interest will result in a collection of strains with disrupted essential genes.

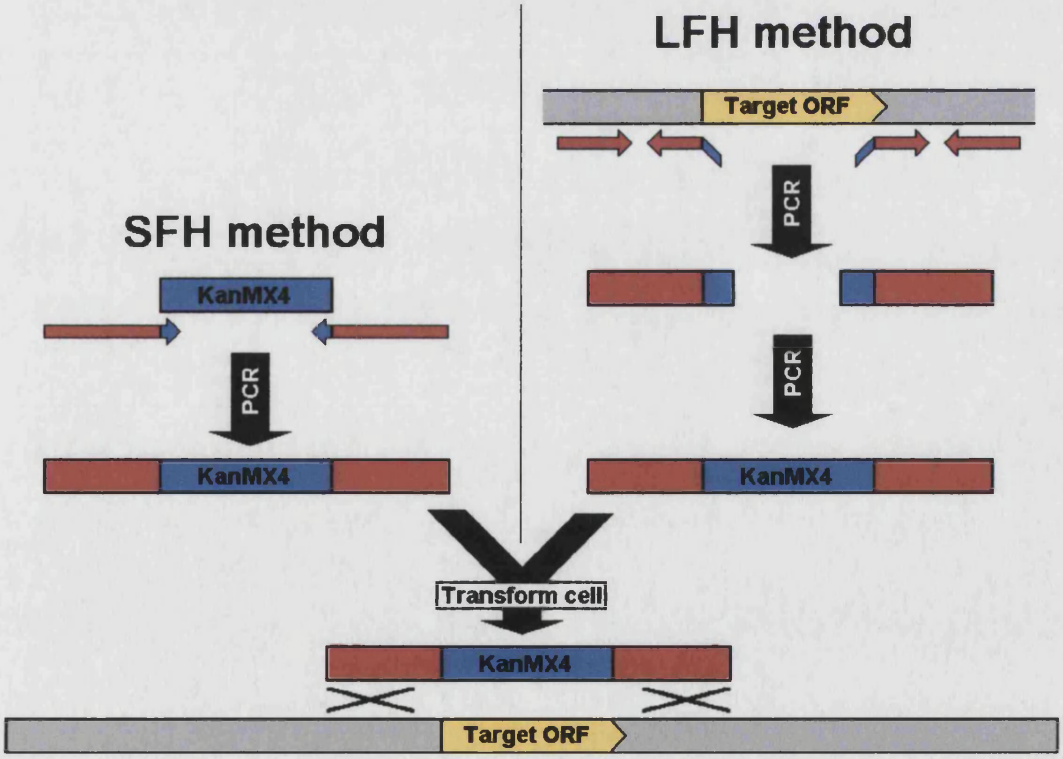


Diagram 1-9: The two methods, Short-Flanking Homology (SFH) and Long-Flanking Homology (LFH), used to create the deletion cassette for the deletion of a target gene for the EUROFAN project.

However after identifying each of the transposed strains, the gene that has been disrupted must be identified. Inevitably the same gene will be isolated many times. If very few genes are involved in the process of interest or the inserts are non-random a very large number of genes must be tested before a reasonable degree of certainty that all/virtually all the genes in the genome have been tested at least once. It has been estimated that in excess of 40,000 transposon disruption strains need to be tested in order to isolate mutants of 90% of the genes in the genome (Ross-Macdonald, Coelho et al. 1999).

Gene knockouts for the *Saccharomyces* Genome Deletion Project (SGDP) were created in a very similar way to those created for the EUROFAN project. The aim of the SGDP was to generate as complete a set as possible of yeast deletion strains with the overall goal of assigning function to the ORFs through phenotypic analysis of the mutants (http://www-sequence.stanford.edu/group/yeast_deletion_project/deletions3.html). A PCR-based strategy was used to create a deletion cassette in a similar way as the gene deletions for the EUROFAN project. Transformation and homologous recombination was used to replace the gene with the *kanMX4* marker (Diagram 1-10) but during the PCR stage a unique sequence was added to the regions flanking the marker gene. These sequences can be used to confirm the presence/identity of the gene knockout should it become necessary (http://www-sequence.stanford.edu/group/yeast_deletion_project/PCR_strategy.html).

Bioinformatics

In recent years the development of high-throughput screening techniques has led an explosion in the amount of data available to yeast researchers. While all this information is very useful, the large volume of data can make it difficult to find specific information and identify patterns. Consequently bioinformatic techniques have evolved to aid the researcher in sorting and using the information available to them. The Internet is instrumental in this, by providing easy and quick access to large amounts of data and programs with which to interpret it on a global scale.

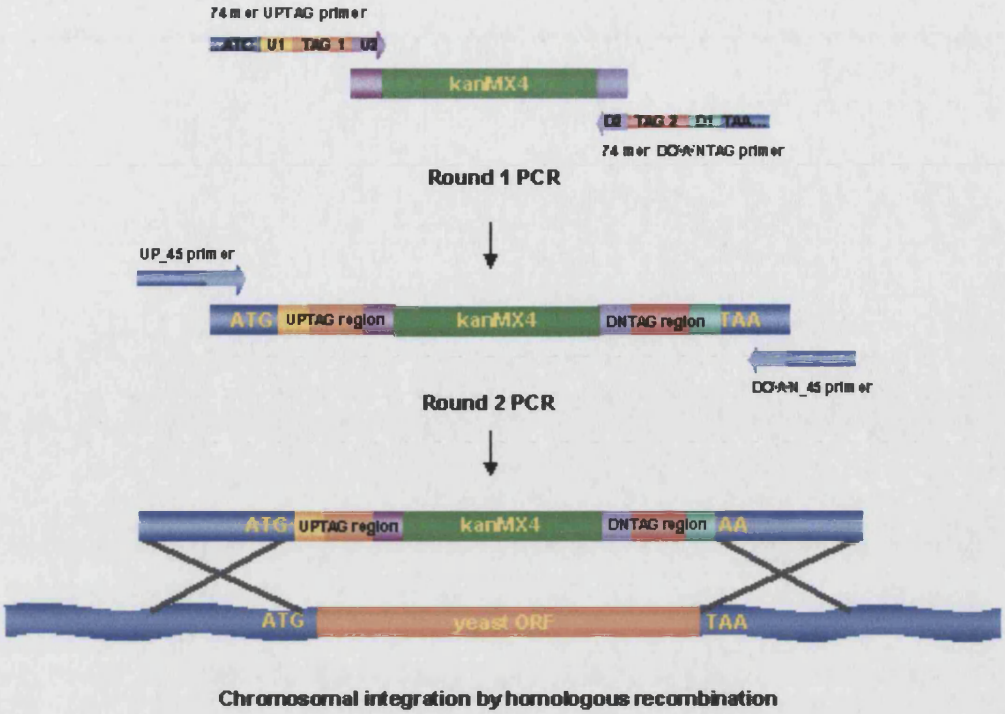


Diagram 1-10: Method used to create the gene knockouts for the *Saccharomyces* genome deletion project (PCRstrat.gif from http://www-sequence.stanford.edu/group/yeast_deletion_project/PCR_strategy.html).

There are many databases of interest to yeast researchers available over the Internet (Table 1-1), and many more of a more general biological and scientific focus (Walsh and Barrell 1996; Baxevanis 2000; Henstock and Wheals 2002). They range from general genome and proteome databases (e.g. EMBL (www.ebi.ac.uk/embl/) and NCBI (http://www.ncbi.nlm.nih.gov/)) to more specialised databases (such as organelle and metabolic databases (Gerrard, Sparrow et al. 2001; Wolfsberg, Schafer et al. 2001)).

With a very large volume of information manual analysis of the data becomes impossible. Consequently there are many software programs available to analyse the data. For DNA sequences there are programs to aid in the construction and annotation of genomes from raw sequencing data. While programs such as BLAST and FASTA are used

to search sequence databases for sequences similar to known gene sequences or potential gene sequences identified during annotation of a DNA sequence.

BLAST and FASTA can also be used to search for homology using protein sequences. It is better to search for homologous gene sequences using protein sequences rather than DNA sequences because protein sequences are less degenerate than DNA sequences and thus the number of misleading and poor matches will be reduced. For genes of unknown function clues to the function of the protein they encode can be obtained from the protein sequence. Programs such as BLOCKS (http://www.blocks.fhcrc.org) and PRINTS (http://www.bioinf.man.ac.uk/dbbrowser/PRINTS) search for sequence motifs that characterise families of proteins. Others programs can, using methods such as energy minimisation, give a prediction as to how the protein will fold. The accuracy of these predictions can be dramatically improved if the structure of a protein with a similar sequence is known.

Database	URL	Description
MIPS	http://mips.gsf.de	Nucleotide and protein information database. Also hosts EUROFAN data, and other useful pieces of information.
SGD	http://www.yeastgenome.org/	Nucleotide and protein database. Also has a wide variety of other information and datasets.
YPD	www.proteome.com/YPDhome.html	Very comprehensive protein database, but is not free.
SCPD	http://cgsigma.cshl.org/jian/	Database of promoter sequences from *S. cerevisiae.*
Gene Ontology	http://www.geneontology.org	Nucleotide and protein database, which provides data on a gene/protein and its orthologs in different species.

Table 1-1: Some examples of different databases available on the internet

With the advent of DNA microarray techniques it has become possible to examine the transcriptional activity of every gene in a genome simultaneously and this has led to an extremely large amount of data being generated. Once data has been collected for a number of time points/experimental conditions, programs such as Cluster (Eisen, Spellman et al. 1998) can be used to identify those genes that are being regulated in a similar manner (thus suggesting they are involved in the same/similar processes). There are also programs (e.g. Treeview (Eisen, Spellman et al. 1998), which uses the output of the Cluster program) that can take the numerical microarray data and convert it into a graphical format making the data easier to view.

All the bioinformatics/data analysis applications now available to researchers can be a set of powerful tools. However if the data are of poor quality or wrong, or incorrect settings are used, then the predictions generated will be inaccurate or misleading. Once a prediction has been made "wet" laboratory experiments should be conducted to confirm it.

Internal Ribosome Entry Sites (IRES)

Translation of eukaryotic mRNAs occurs mainly using the ribosome scanning mechanism in which the 40S subunit of the ribosome, and associated initiation factors, attaches to the 5'-CAP structure of the mRNA and scans along the mRNA until it reaches the start codon of the ORF. At this point the 60S subunit associated with the translation complex and the initiation factors dissociate and the translation of the ORF starts (Diagram 1-11) (reviewed in (McCarthy 1998; Kozak 1999; Sachs and Varani 2000). However internal transcription initiation has also been observed. In this mechanism, a ribosome is recruited and directed to the start codon independently of the 5'-CAP structure. The site that allows this to occur is known as an Internal Ribosome Entry Site or IRES. This is similar to the mechanism of translation initiation in prokaryotes, where the Shine-Dalgarno sequence recruits the ribosome to the prokaryotic mRNA (Kozak 1999).

IRES are known to exist in mammalian and viral systems (Kozak 1999; Martínez-Salas, Ramos et al. 2001; Pestova, Kolupaeva et al. 2001). During picornavirus infection the majority of translation initiated using an IRES element, as this bypasses CAP-dependent translation inhibition caused by a very low level of eIF4F (Martínez-Salas, Ramos et al. 2001). Comparing IRES from picornaviruses, and IRES from distantly related viral and cellular systems, indicates that both IRES sequence and length are not conserved. While phylogenetic conservation of secondary structure suggests that the shape of the IRES is an important factor (Martínez-Salas, Ramos et al. 2001). This has prevented either the delineation of a consensus sequence or the construction of a model on how IRES sequences might work (Martínez-Salas, Ramos et al. 2001; Pestova, Kolupaeva et al. 2001).

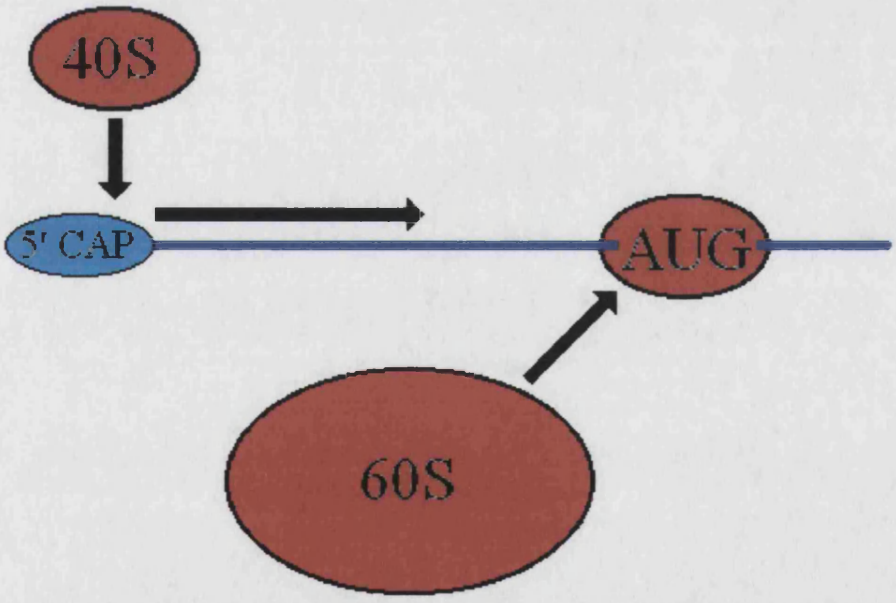


Diagram 1-11: Translation initiation in *S. cerevisiae*. The 40S subunit and associated factors bind to the 5'-CAP structure and scan along to the start codon. At this point the initiation factors dissociate, the 60S subunit associates and translation of the ORF starts.

Research conducted by a number of laboratories (Tranque, Hu et al. 1998; Hu, Tranque et al. 1999; Zhou, Edelman et al. 2001) has shown that by analogy with prokaryotic translation initiation (Pestova, Kolupaeva et al. 2001) homology of the 5'-UTR of the mRNA with the 3' end of the 18S rRNA may be significant in internal translation initiation. Until recently functional IRES have not been shown to occur in yeast cells, although yeast cell extracts can support IRES mediated translation initiation (Martínez-Salas, Ramos et al. 2001). In 1999 an *Escherichia coli* sequence was found to act as an IRES in starved *S. cerevisiae* cells (Diagram 1-12A) (Paz, Abramovitz et al. 1999). Certain stimuli, including serum deprivation and growth arrest (Martínez-Salas, Ramos et al. 2001) are known to activate known cellular IRES. This suggests that *S. cerevisiae* is using IRES in a similar manner to its eukaryotic counterparts. More recently the 5'-UTR of the *S. cerevisiae* genes YAP1 and p150 has been shown to contain IRES that are functional in vegetatively growing cells (Diagram 1-12B) (Zhou, Edelman et al. 2001).

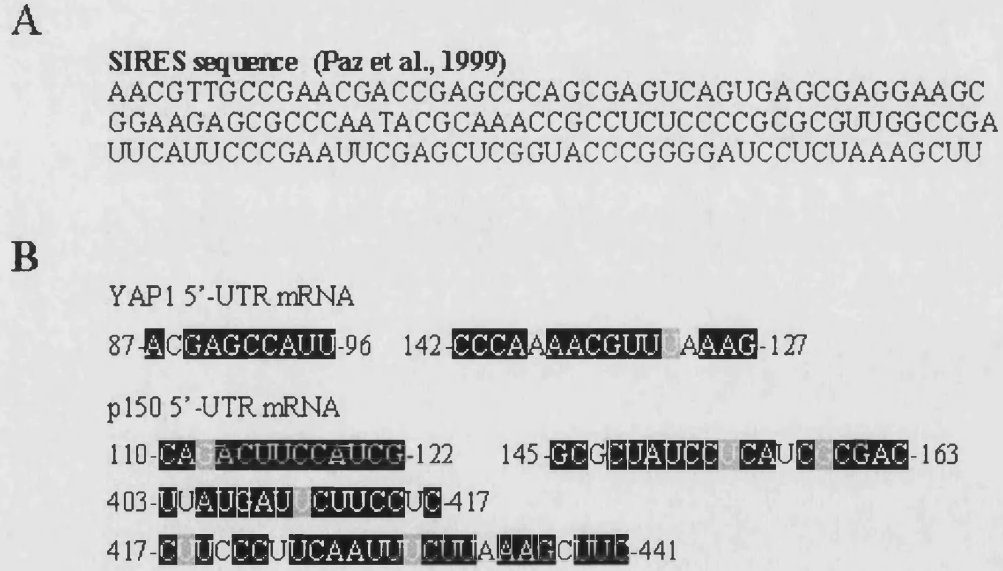


Diagram 1-12: Sequences known to aid in the internal initiation of translation in *S. cerevisiae*. A: The SIRES sequence (Paz, Abramovitz et al. 1999). B: Sequences from the 5'-UTR of the YAP1 and p150 mRNAs which are complementary to the 18S rRNA. Nucleotides shaded in black denote base pairing, and grey shading indicates GU base-pairing (Zhou, Edelman et al. 2001).

Aims and objectives of this research

The development of DNA microarrays has allowed the large-scale analysis of the transcriptome of yeast cells. Several studies have looked at the yeast transcriptome under many conditions including stationary phase (DeRisi, Iyer et al. 1997; Chu, DeRisi et al. 1998; Eisen, Spellman et al. 1998; Spellman, Sherlock et al. 1998; Ferea, Botstein et al. 1999; Gasch, Spellman et al. 2000; Gasch, Huang et al. 2001). While this provides very useful information on which genes are upregulated during stationary phase, and are therefore probably required for stationary phase, it doesn't indicate those genes that are essential for stationary phase. The aim of this investigation was to study stationary phase and reveal the components and processes that are essential for it. The first phase of the study used single gene knockout strains generated by the EUROFAN2 and *Saccharomyces* Genome Deletion Projects. By growing these strains into stationary phase it should be possible to identify those genes that provide a function that is essential for the maintenance of viability in stationary phase. The second phase of the study used those genes identified in the first phase to answer questions about stationary phase. For example, do the functions of these genes reveal anything about the processes and functions required for stationary phase? What causes the gene knockouts to lose viability in stationary phase? How does this relate to the essentiality of the functions? The essential genes can also be used to investigate the transcriptional regulation of stationary phase essential genes.

Chapter 2

Screening for stationary phase essential genes

Summary

ORF knockout sets were grown to stationary phase and left to incubate for approximately three months. Approximately 100 ORFs were found to be essential for maintenance of cell viability in stationary phase after this length of time. These ORFs revealed that genes of a wide variety of functions were essential for surviving stationary phase. A large proportion of these genes was related to respiratory and/or mitochondrial functions, especially mitochondrial ribosome functions. Retests of the identified genes and other sets of genes using different media revealed that some of the identified genes were not essential under all conditions. The ability to respire was not essential for maintaining viability in stationary phase.

Stationary phase essential genes

There have been many different studies of stationary phase in *S. cerevisiae*, some of which have identified genes or proteins that are associated with stationary phase: mutation of some proteins, such as Ard1p, eliminates the ability of a cell to react fully to stationary phase; Ypt1p is essential during nitrogen, but not carbon, starvation; Hsp104p mutants have a marginally reduced ability to maintain viability in stationary phase; the *ils1-1* mutation causes a cell to acquire stationary phase characteristics under non-starvation conditions (Werner-Washburne, Braun et al. 1993). However the contribution of the majority of the proteins in the *S. cerevisiae* proteome remains unknown.

While there have been large-scale studies on the transcriptome of stationary phase cells there has not been a systematic study of gene knockouts to find genes essential for the

maintenance of viability during long term stationary phase. In any process there are genes with a function that is essential. By growing viable ORF knockout strains into stationary phase it may be possible to screen for those genes/proteins that have a function that is essential for stationary phase - absence of the essential function would lead to inability to enter or survive or exit from stationary phase and lead to the death of the cell.

Materials and Methods

Strains used

Two sets of targeted gene knockouts were used: the EUROFAN set (http://mips.gsf.de/proj/yeast/CYGD/db/index.html) and the *Saccharomyces* Genome Deletion Project set (SGDP) (http://sequence-www.stanford.edu/group/yeast_deletion _project/deletions3.html). Both sets of gene knockouts were constructed in *S. cerevisiae* FY1679 (*a/alpha ura3-52/ura3-52 leu2Δ1/+ trp1Δ63/+ his3Δ200/+*) (Winston, Dollard et al. 1995).

Media Used

YPD 1 % (w/v) yeast extract, 2 % (w/v) peptone, and 2 % (w/v) dextrose (plus 2 % (w/v) agar if solid medium was required) was dissolved in water and autoclaved to sterilise (Brown and Tuite 1998).

YP(0.3D) YPD medium with 0.3% (w/v) dextrose instead of 2% (w/v) dextrose.

SC 6.7 g yeast nitrogen base (without amino acids), 400 mg uracil, 10 mg adenine, 50 mg tyrosine, 20 g dextrose, 10 ml 0.5 M disodium hydrogen phosphate solution was dissolved in 990 ml distilled water and autoclaved to sterilise. After autoclaving 10 ml of sterile amino acid solution was added (amino acid solution is made from 200 mg arginine, 200 mg histidine, 600 mg isoleucine, 600 mg leucine, 400 mg lysine, 100 mg methionine, 600 mg phenylalanine, 1000 mg threonine, 400 mg tryptophan dissolved in water to

a final volume of 100 ml and filter sterilised using a 0.22 μm filter) (Brown and Tuite 1998).

SC(0.3D) SC medium with 3g of dextrose instead of 20g dextrose.

YPGlycerol 1 % (w/v) yeast extract, 2 % (w/v) peptone, and 2 % (v/v) glycerol (plus 2 % (w/v) agar if solid medium is required) was dissolved in water and autoclaved to sterilise (based on YPDG medium (Brown and Tuite 1998)).

YPGalactose 1 % (w/v) yeast extract, 2% (w/v) peptone, and 2 % (w/v) galactose (plus 2 % (w/v) agar if solid medium is required) was dissolved in water and autoclaved to sterilise (based on YPD medium (Brown and Tuite 1998)).

Testing growth arrest is due to glucose starvation

Eight 5 ml Falcon tubes each containing 2.5 ml liquid YPD medium were inoculated with *S. cerevisiae* FY1679. Twice daily a sample of the culture was removed and observed under a microscope. The ratio of budded to unbudded cells in the medium (budding index) of each sample was recorded. When the budding index appeared to have stopped dropping and become stable, 0.5 ml of 10% dextrose solution was added to four of the tubes. To the other four tubes 0.5 ml of sterile distilled water was added. This volume of dextrose solution was added to the medium to return the dextrose concentration in the medium to the original concentration of 2%. The water was added to the other tubes to match the change in culture volume. Following the addition of dextrose/water the budding index of the cultures was monitored for any changes.

Testing for stationary phase defects

Knockout strains were inoculated into 2.5 ml of liquid YPD media in 5 ml Falcon tubes. These were allowed to grow for 7 days under aerobic conditions at 28°C with shaking. At this point the tubes were examined and it was noted if the culture had grown. The cultures were then sealed and left at 28°C until 93 days post-inoculation, at which time

3 µl samples were taken and spotted onto solid YPD medium. The samples were allowed to grow for 30 hours at 28°C, after which the growth of the sample was scored (Diagram 2-1).

Repeats were undertaken on only those gene knockouts that showed a severe loss in viability after 93 days in the MAT**a**, MATα, and homozygous diploid strains, but not in the heterozygous diploid strain. For the repeats of the potentially stationary phase defective gene knockouts the cells were grown in 2.5 ml of liquid YPD, YP(0.3D), SC, SC(0.3% D), YPGlycerol, YPGalactose, and YPD with GSH (YPD medium with 20 µl 0.125M reduced glutathione (GSH) added to each culture at 1 week post-inoculation. This gives a final concentration of 1 mM glutathione in the culture). The viability of the cultures was tested after 93 days by observing growth of a sample on solid YPD medium.

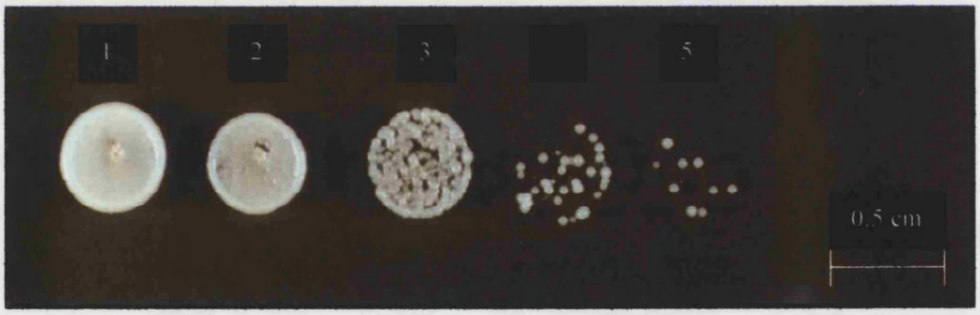


Diagram 2-1: A dilution series showing yeast growth. Spots 1 and 2 show confluent growth and spot 3 shows "grainy" growth. A culture of wild type cells, grown to a cell density of approximately 1 x 10^8 cells.ml^{-1} would produce a spot similar to spot 1. A countable number of colonies have grown in spots 4 and 5, this level of growth after 90 days would be taken as an indication of a severe stationary phase defect. Spot 4 (32 colonies) and spot 5 (10 colonies), if grown from a 3 µl sample, would indicate 1 x 10^4 and 3.3 x 10^3 viable cell.ml^{-1} (to 1 sig. fig.) respectively.

Monitoring loss of cell viability in the wild type strain

S. cerevisiae FY1679 was inoculated into 50 ml of YPD medium and was grown at 28°C, with shaking, for one week. The culture was split into 13 x 2.5 ml aliquots and put into 4.5 ml Falcon tubes. The Falcon tubes were sealed shut and stored static at 28°C. Every week, from 1 week to 13 weeks post-inoculation, the viability of the culture was determined using the culture from a different tube each week. On the first week 10^{-4}, 10^{-5} and 10^{-6}-fold dilutions of the culture was plated out onto YPD medium and grown for two days at 28°C. The number of colonies growing was then used to estimate the number of viable cells in the culture. After the first week only one dilution, based on the results of the previous week, was used.

Results

Glucose induced growth arrest

The budding index of eight cultures, named A to H, of *S. cerevisiae* FY1679 was followed for seven days. At this point the budding index of all the cultures appeared to have become stable. Water was added to cultures A to D and glucose to cultures E to H and the budding index was followed for three more days (Diagram 2-2). The high budding index shows that immediately after inoculation *S. cerevisiae* FY1679 cultures were growing at a fast rate. However after about two to three days the percentage of budded cells in the cultures fell to approximately 23% and remained at this level for the next four to five days indicating that the cultures had entered a state of growth arrest.

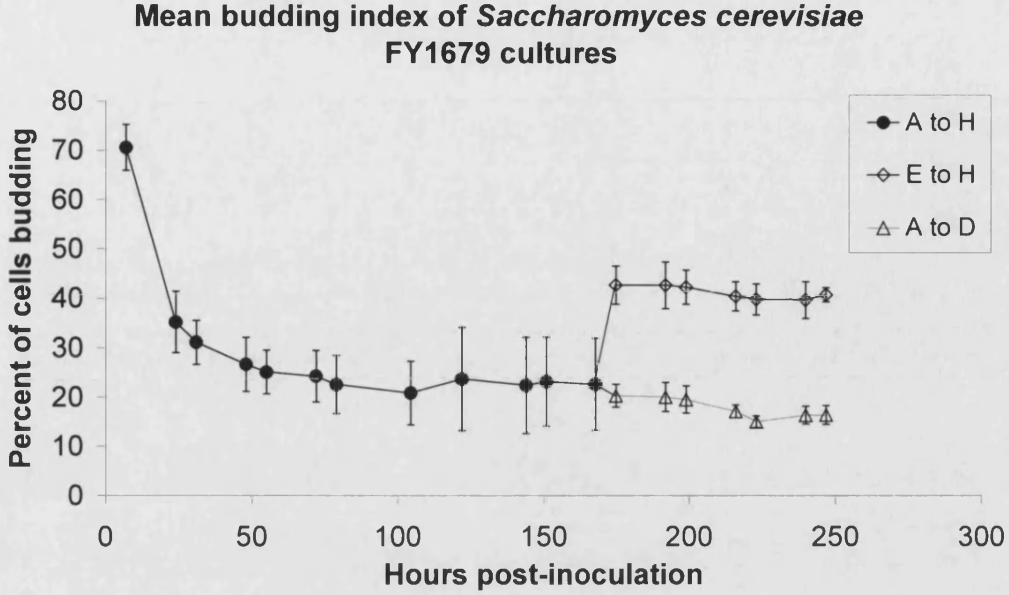


Diagram 2-2: The budding index of eight cultures of *S. cerevisiae* FY1679 in YPD medium. At 168 hours after inoculation water was added to cultures A to D, and glucose solution to cultures E to H. Each datapoint is the mean ± Standard Error.

Addition of water led to no significant change in budding index as it fell slowly to approximately 20%. Addition of glucose led to an immediate increase in budding index that was maintained at 40%. This indicates that the growth arrest that the cultures had entered previously was due to glucose starvation and was not due to the change in culture volume as the addition of glucose stimulated growth but the addition of water did not. The budding index did not reach a level as high as seen immediately after inoculation probably because addition of glucose alleviated carbon starvation but the cells quickly become starved for some other nutrient, such as nitrogen, as the only component of the medium that was refreshed was the carbon source.

Loss of wild type strain cell viability

The number of viable cells in the culture of *S. cerevisiae* FY1679 each week was calculated using the formula:

$$\text{Number of viable cells per ml of culture} = \frac{\text{Number of colonies grown on plate}}{\text{Volume (µl) of diluted clture used}} \times \frac{1000}{\text{Dilution Factor}}$$

The calculated numbers of viable cells and the loss of viability is given in Table 2-1 and diagram 2-3.

Week	1	2	3	4	5	6	7
viable cells per ml	7.90×10^7	1.04×10^8	1.41×10^8	8.15×10^7	4.75×10^7	2.30×10^7	2.50×10^6
Week	8	9	10	11	12	13	
viable cells per ml	3.70×10^6	2.40×10^7	1.30×10^6	2.50×10^6	2.00×10^6	8.50×10^5	

Table 2-1: The loss of viability of a culture of *S. cerevisiae* FY1679

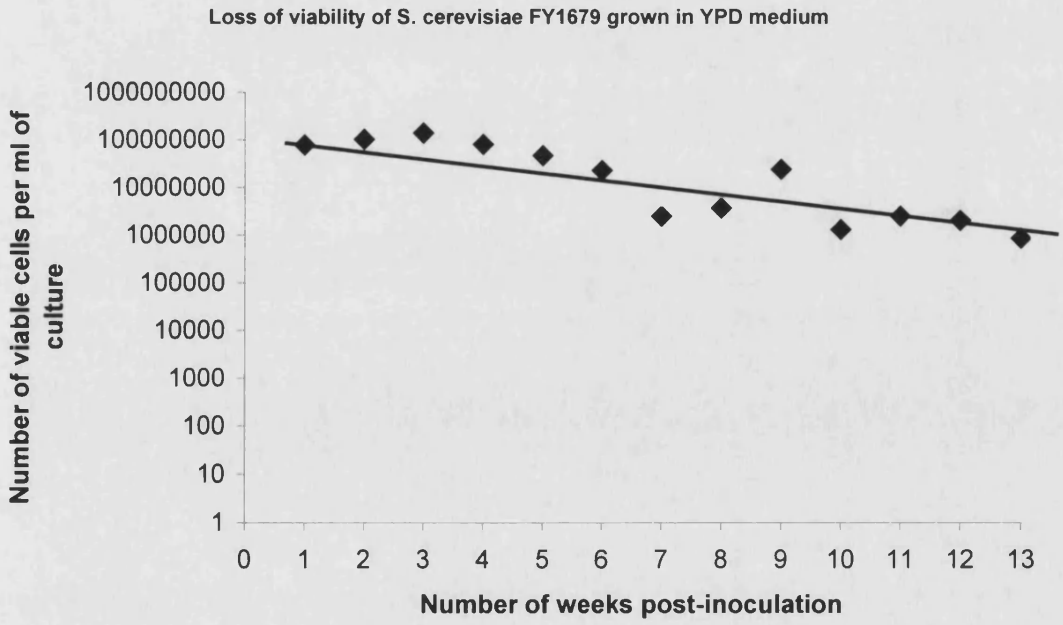


Diagram 2-3: A semi-log plot of the loss of viability of a culture of *S. cerevisiae* FY1679. Black diamonds represent actual data and the linear black line was fitted by eye.

The culture of *S. cerevisiae* shows an approximately 100-fold drop in viability. From about $8x10^7$ to $8x10^5$ viable cells per ml of culture. On a semi-log plot the loss of viability appears to occur in a linear fashion. This would suggest that the loss of viability in this culture is following an exponential decay which can be interpreted as a constant probability of a cell dying per unit time.

The value of $8x10^5$ viable cells per ml of culture at the end of the experiment still represents a relatively large number of viable cells. If a 4µl sample was grown on solid medium this concentration of viable cells would give rise to 3200 colonies which in the spotting method used in this study would produce a confluent spot. Thus the loss in viability exhibited by *S. cerevisiae* FY1679 could not be detected using the spotting method.

To be classified as having a stationary phase defect it was decided that a knockout strain would have to grow less than 50 colonies from a 4µl sample of a stationary phase culture. This represents an approximately 100-fold reduction in viability compared to the wild type strain or approximately a 10^5 reduction in viability overall. This leads to the question do the strains selected represent a subset of strains that have notably reduced viability from the wild type strain or are they simply by chance the first of all the strains to lose viability during stationary phase?

If the strains represent a subset of genes, the loss of which affects stationary phase viability, a graph plotting survival rate against the number of strains would show a distinct peak at low levels of viability. On the other hand, if there is no distinct group of stationary phase essential genes the numbers of strains will increase smoothly as survival increases in a plot of viability versus number of strains.

The results showed (diagram 2-4) a peak in the number of strains at very low levels of viability while most of the samples exhibit grainy or more confluent growth. This result suggests that the genes identified represent a subset of gene knockouts that

affect stationary phase viability to a greater extent than either other gene knockouts or the wild type strain.

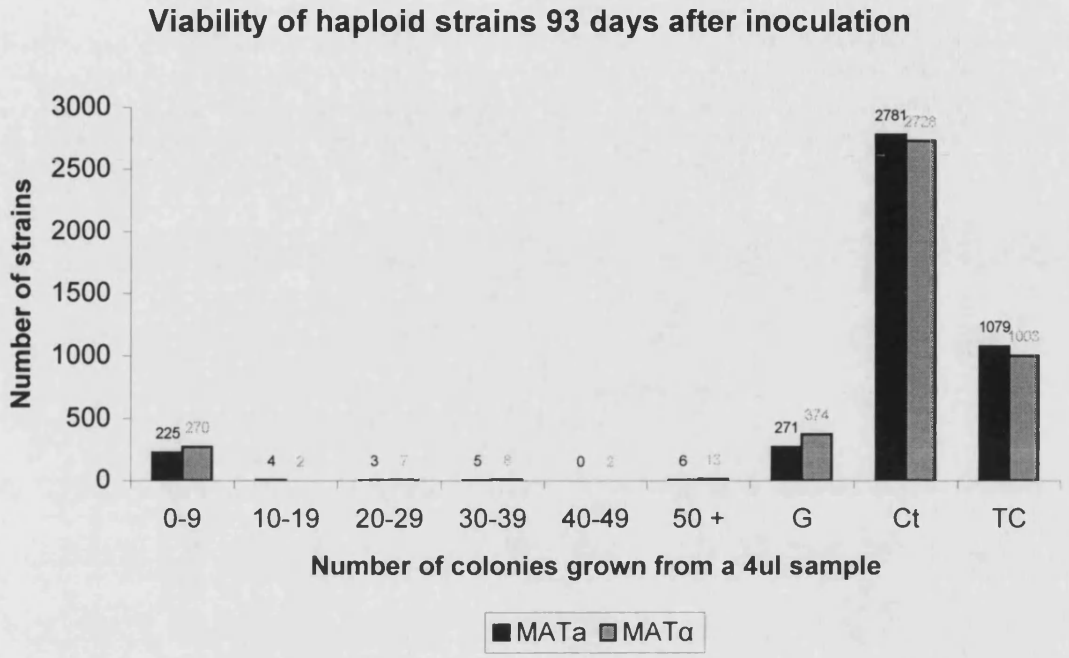


Diagram 2-4: A graph showing the frequency of haploid strains (from the SGDP set of ORF knockouts) growing specific numbers of colonies from 4µl samples of stationary phase cultures. The letters G, Ct and TC indicate samples that have grainy, thin confluent or thick confluent growth respectively. The numbers above the bars indicate the number of strains involved.

Screening EUROFAN ORF knockouts

To be classified as having a stationary phase defect the diploid strain must have showed no significant loss in viability and both haploid strains produced ≤ 100 colonies (a loss in viability of at least 10^6-fold) when spotted out after 90 days. Those strains where the diploid and both haploid strains showed reduced viability were classified as having a dominant stationary phase defect.

Of the 768 ORF knockouts tested from the EUROFAN set of gene knockouts, 95 ORF knockouts (12.4%) were classified as having a stationary phase defect after the first screen (appendix 3). Cell wall structure has also been shown to be required for stationary phase survival (McGraw and Henry 1989; Werner-Washburne, Braun et al. 1993). Defects in stationary phase might be expected also to cause defects in sporulation and mating, as these processes are also induced by starvation conditions. Over half of these knockouts with a stationary phase defect had either no other defect, or a defect in one of these closely related processes (Table 2-2).

Those genes with defects in stationary phase and processes other than sporulation, mating and/or the cell wall are less likely to be genes that are directly involved in stationary phase. The lethal phenotype that is observed is more likely, in these cases, to be a secondary consequence of the ORF knockout. In other words the knockout strains die in stationary phase not because part of the process used in stationary phase is not functioning but because the ORF knockout is causing the cell to become very unfit.

Further work was not conducted on this set of ORF knockouts for several reasons. Compared to the ORF knockouts from the *Saccharomyces* Genome Deletion Project this set of ORF knockouts showed poor congruence between the different haploid mating types of the same ORF knockout. This lack of similarity is suspicious because unless the knocked out ORF encodes a protein used in mating or sex type determination these two results should be he same. The results also displayed a far greater range of loss of viability across the whole screen. The range of levels of growth from the EUROFAN set of knockouts means that there is no definitive point that can be used to separate strains that do and do not survive. With the SGDP set of ORF knockouts the strains generally did not grow at all or produced confluence after 93 days. This can be seen in diagram 2-4.

Phenotype	Number of ORFs
Stationary phase	11
Stationary phase and sporulation	5
Stationary phase and mating	3
Stationary phase and cell wall	8
Stationary phase, sporulation, and mating	3
Stationary phase, sporulation, and cell wall	2
Stationary phase, sporulation, mating and cell wall	1
Stationary phase and other phenotypes but not sporulation, mating or cell wall	0
Stationary phase, other phenotypes, and sporulation, mating, and/or cell wall	25
Total	58

Table 2-2: Number of genes with stationary phase and/or sporulation/mating/cell wall defects.

In addition to this the behaviour of the EUROFAN set of ORF knockouts was not consistent in subsequent analyses of this screen and in experiments by other researchers (Alain Nicolas, personal communication). This means that the results of the screen cannot be trusted as they are not repeatable. There was also a lack of congruence between the two sets of ORF knockouts in the genes identified as having a stationary phase defect. Most genes that were identified as having a stationary phase defect from the EUROFAN set of ORF knockouts were not identified as having a stationary phase defect when they were screened as part of the SGDP set. As the method to knockout the ORFs for each of the projects was similar and the method used in the screen for a stationary phase defect was identical, ORFs that appear in both ORF knockout sets should give identical results. One

significant difference between the construction of the two sets was that the EUROFAN set was constructed by about 150 European labs, many of whom were inexperienced in the methodology whereas all the SGDP sets were constructed by just 16 very experienced laboratories.

Screening *Saccharomyces* Genome Deletion Project ORF knockouts

Of the 4424 ORF knockouts from the SGDP set for which cultures of both haploid strains were obtained, 122 ORF knockouts were classified as having a stationary phase defect (growth of less than 50 colonies from a 4 μl sample) after the first screen (appendix 3). One or both of the haploid strains were not available, did not grow, or were lost due to contamination for another 1459 ORF knockouts. Where results of one of the two haploid strains could not be obtained (e.g. absence of a strain, contamination) the ORF knockout was not classified as causing a stationary phase defect. These ORF knockouts were retested using YPD medium, as in the primary screen, to confirm the stationary phase defective phenotype.

They were also tested in various other media to test the effect of various conditions on the stationary phase defects of the ORF knockouts. In addition to YPD the strains were retested in YPD with 0.3% instead of 2% dextrose, YP(0.3D). Recent work (Lin, Kaeberlein et al. 2002) has shown that for genealogical ageing, increased respiration caused by calorie restriction (C.R.) (growth in YPD with 0.5% dextrose) will increase the life span of *S. cerevisiae*. YPD is a rich medium containing many different biological compounds. The knockout strains were also tested in SC and SC(0.3D) (SC medium with 0.3% instead of 2% dextrose) media, as these are less rich than YPD medium.

Many studies into ageing in *S. cerevisiae* have implicated oxidative damage as a cause of ageing (Jakubowski, Bilinski et al. 2000; Nestlbacher, Laun et al. 2000). To reduce any impact of oxidative damage, reduced glutathione (GSH) was added to the growth medium to a final concentration of 1 mM. Reduced glutathione, at a concentration

of 1 mM in the growth medium, protects yeast against oxidative damage (Grant, MacIver et al. 1996; Nestlbacher, Laun et al. 2000). The GSH was added to the cultures at 1 week post-inoculation. If the GSH was added at the point of inoculation it would have be completely oxidized by the end of the week of aerobic growth. It would then offer no protection to the cells against oxidative damage while the cells were in stationary phase.

When the 122 ORF knockouts were retested 30 were shown repeatedly not to lose viability after 93 days in YPD (overview Table 2-3, appendix 1 (highlighted strains) and appendix 3), leaving a total of 92 ORF knockouts that have been shown to lose viability after 93 days in stationary phase. Of the 30 strains that were found to not have a stationary phase defect when retested in YPD medium, 17 were found to have no defect when grown in any of the media used in this investigation. The other 13 strains had a stationary phase defect when grown in SC medium (appendix 3), with three of these strains (knockouts of the ORFs YDL067c, YDR197w, and YER155c) also having a stationary phase defect when grown in YPGalactose medium. The results of all these retests are shown in appendix 1.

The relatively high proportion (approximately 25%) of strains identified in the primary screen but rejected on retesting was probably due to the manner in which the screen was run. Due to the limited amount of space and time only the two haploid strains for each ORF knockout were screened with no repetition. This meant that it was impossible to identify those results at that stage that were anomalous and should therefore be ignored. This would also suggest that there are a number of ORF knockouts that may have been incorrectly identified as being non-essential for stationary phase in the primary screen. The final set is thus likely to be less than the total number that might have a severe defect. However, this limitation is true for all screens that will miss some potential candidate strains.

Nº of genes	Defect in YPD medium				Defect in SC medium		Other media		
	2% dextrose	0.3% dextrose	With GSH	With Sorbitol	2% dextrose	0.3% dextrose	YPG	YPGal	resp. comp.
Yes	87	9	19	47	88	73	11	39	60
No	30	108	80	57	26	45	45	64	62
?	5	5	13	18	8	4	4	19	0
-	0	0	0	0	0	0	62	0	0
Total	122	122	112	122	122	122	122	122	122

Table 2-3: Overview of the results of testing ORF knockouts for stationary phase defects in various different media. Respiratory competency was assumed if the knockout strain was able to grown on YPG medium. "Yes" and "No" indicate respectively whether an ORF knockout strain does or does not have a stationary phase defect in that medium (a loss of viability after three months in spent medium). "?" and "-" respectively indicate that no conclusion could be drawn or there are no results. Respiratory competency ("resp. comp." Column) is defined as the ability ("Yes") or inability ("No") to grow on YPG medium.

Stationary phase essential ORF knockouts

A large amount of relevant data about these genes can be found on remote databases (Henstock and Wheals 2001; Henstock and Wheals 2002). By collating these data it is possible to gain a better insight into any possible function of a gene. The information collected about each of the ORFs enables the ORFs to be grouped together by the function of the proteins for which the genes encode (Table 2-4).

	ORF	Other names	Description
Unknown Function	YDR065w	YD9609.19, D4258	Hypothetical ORF
	YGR102c	G5930	Hypothetical ORF
	YGR150c	G6642	Hypothetical ORF
	YMR098c	YM6543.05	Hypothetical ORF
	YOR305w	O5653	Hypothetical ORF
	YPR099c	P8283.13A	Hypothetical ORF
	YPR100w	P8283.12	Weak similarity to *C. elegans* hypothetical protein CEC25A1
	YPR116w	P8283.2	Hypothetical ORF
	YDL068w	D2518	Hypothetical ORF
	YGR160w	G7004	Hypothetical ORF
	YNR036c	N3298	Strong similarity to ribosomal protein S12
	YGL246c	RAI1, NRE387, G0580	Weak similarity to *C. elegans* dom-3 protein
transcription and RNA maturation	YFL036w	RPO41	Mitochondrial DNA-directed RNA polymerase
	YNL139c	RLR1, THO2, ZRG13, N1835	Plays a role in transcription elongation by RNA polymerase II
	YJR122w	CAF17, J2043	CCR4 transcriptional complex component
	YMR228w	MTF1, mtTFB, RF1023, YM9959.10	Mitochondrial RNA polymerase specificity factor
	YMR282c	AEP2, ATP13, YM8021.08	Required for the expression of subunit-9 of ATP synthase, weak similarity to *S. pombe* rad3
	YDL107w	MSS2, D2340	COX1 pre-mRNA splicing factor
	YHL038c	CBP2	Apo-cytochrome b pre-mRNA processing protein 2 (required for splicing of COB bI5 intron).
	YLR203c	MSS51, L8167.17	Protein involved in maturation of COX1 and COB mRNA.

Table 2-4 (Part 1): Genes identified as being essential for stationary phase (information collected from the MIPS and SGD databases.

	ORF	Other names	Description
(transcription and RNA maturation continued)	YMR064w	AEP1, NCA1, YM9916.03	Nuclear control of ATPase mRNA expression protein, required for accumulation of mitochondrial transcript of ATP9/OLI1
	YDL044c	MTF2, NAM1, D2705	Mitochondrial protein involved in mRNA splicing and protein synthesis
	YIR021w	MRS1, PET157	RNA splicing protein of the mitochondrial carrier (MCF) family.
	YDR194c	MSS116, YD9346.05	Mitochondrial RNA helicase of the DEAD box family
	YKR024c	DBP7	RNA helicase required for 60S ribosomal subunit assembly
	YPL029w	SUV3, LPB2	Mitochondrial ATP-dependent RNA helicase
translation and protein maturation	YLR067c	PET309, L2189	Required for stability and translation of COX1 mRNA
	YDR197w	CBS2, CBP7, YD9346.08	Cytochrome b (COB mRNA) translational activator
	YJL102w	MEF2, J0826	Mitochondrial translation elongation factor
	YLR069c	MEF1, L2195	Mitochondrial translation elongation factor
	YGL143c	MRF1, G2530	Mitochondrial polypeptide chain release factor
	YHR038w	FIL1, KIM4, H8179.10	Involved in mitochondrial protein synthesis, regulatory factor involved in glucose repression/ derepression. Similarity to prokaryotic ribosome releasing/recycling factor
	YER154w	OXA1, PET1402, PET-TS1402, HCY69	Cytochrome oxidase biogenesis protein, mediates the export of proteins from the mitochondrial matrix to the intermembrane space.
	YDR518w	EUG1, D9719.23	Protein disulfide isomerase
	YGL135w	RPL1B, SSM1B, SSM2, G2834	60S large subunit ribosomal protein (L1B)

Table 2-4 (Part 2): Genes identified as being essential for stationary phase (information collected from the MIPS and SGD databases.

	ORF	Other names	Description
(translation and protein maturation continued)	**YCR046c**	IMG1, PETCR46	Mitochondrial ribosomal protein, required for respiration and maintenance of mitochondrial genome
	YDL045w-a	MRP10	Yml37p homolog. Mitochondrial ribosome 37S subunit component
	YPR166c	MRP2, P9325.7	14 kDa mitochondrial ribosomal protein, similar to *E. coli* S14 protein.
	YDR175c	YD9395.08	Similarity to *S. pombe* hypothetical protein SPAC2F7.15
	YDR337w	MRPS28, D9651.3	Mitochondrial ribosomal protein MRPS28 (Similar to *E. coli* ribosomal S15 protein).
	YGL129c	RSM23, RSM51, G2856	ATPase (putative). Mitochondrial ribosome small subunit component
	YJR113c	RSM7, J2020	Mitochondrial ribosome small subunit component
	YMR158w	YM8520.07	Weak similarity to *E. coli* ribosomal S8 protein.
	YNR037c	RSM19, N3300	Mitochondrial ribosome small subunit component
	YGR171c	MSM1, G7104	Mitochondrial methionyl-tRNA synthetase
	YPL097w	MSY1, LPG11, SYYM	Tyrosyl-tRNA synthetase
	YPL104w	MSD1, LPG5	Mitochondrial aspartate-tRNA ligase
	YPR047w	MSF1, YP9499.05	Alpha subunit of yeast mitochondrial phenylalanyl-tRNA synthetase
	YPL040c	ISM1, P7102.10	Mitochondrial isoleucine-tRNA ligase
Energy Generation	**YDR298c**	ATP5, OSCP, D9740.11	ATP synthase F_0 sector subunit 5
	YPL078c	ATP4, LPF7	ATP synthase F_0 sector subunit 4
	YPL271w	ATP15, P0345	F_1F_0-ATPase complex, F_1 epsilon subunit.
	YDR148c	KGD2, YD8358.05	2-oxoglutarate dehydrogenase complex E2 component from the mitochondrion
	YKL085w	MDH1, ACN50	Mitochondrial malate dehydrogenase
	YDR178w	SDH4, YD9395.11	Succinate dehydrogenase membrane anchor subunit

Table 2-4 (Part 3): Genes identified as being essential for stationary phase (information collected from the MIPS and SGD databases.

	ORF	Other names	Description
(Energy Generation continued)	YLL041c	SDH2, SDHB, SDH, L0745	Succinate dehydrogenase iron-sulphur protein subunit.
	YDL067c	COX9, D2520	Cytochrome c oxidase subunit VIIa
	YGR062c	COX18, G4532	Required for activity of mitochondrial cytochrome oxidase (cytochrome oxidase gene 18)
	YLL018c-a	COX19	Protein required for cytochrome c oxidase activity
	YML129c	COX14, M_F70, YM4987.06	Cytochrome-c oxidase assembly protein (mitochondrial membrane protein)
	YPL172c	COX10, P2287	Farnesyl transferase required for heme A synthesis (an essential posttranslational stage in assembly of cytochrome oxidase)
	YDR204w	COQ4, YD8142A.01, YD8142.01	Involved in ubiquinone biosynthesis (biosynthesis of coenzyme Q)
	YML110c	COQ5, TCM7, YM8339.09, DBI56	C-methyltransferase (ubiquinone (coenzyme Q) metabolism).
Mitochondrially associated	YBR179c	FZO1, YBR1241	Required for biogenesis of mitochondria. Yeast Fzo homolog (*D. melanogaster* fuzzy onions gene).
	YJR144w	MGM101, MGM9, J2181	Mitochondrial genome maintenance protein
	YHR120w	MSH1	DNA mismatch repair protein, mitochondrial.
	YML061c	PIF1, TST1, YM9958.01	DNA helicase involved in mitochondrial DNA repair and telomere length
	YOL095c	HMI1, HRE571, O0920	Mitochondrial DNA helicase.
	YPR067w	ISA2, YP9499.22	Mitochondrial protein required for iron metabolism (IscA/NifA homolog)
	YLL041c	SDH2, SDHB, SDH, L0745	Succinate dehydrogenase iron-sulphur protein subunit.
	YMR150c	IMP1, PET2858, YM9375.20	Mitochondrial inner membrane protease
	YMR267w	PPA2, IPP2, YM8156.09	Mitochondrial inorganic pyrophosphatase.

Table 2-4 (Part 4): Genes identified as being essential for stationary phase (information collected from the MIPS and SGD databases.

	ORF	Other names	Description
Cell Signalling	YDR507c	GIN4, ERC47, CLA6, D9719.13	ser/thr protein kinase
	YDR523c	SPS1, D9719.27	ser/thr protein kinase
	YPL031c	PHO85, SSG3, P7102.18A	Cyclin-dependent protein kinase. Involved in phosphate and glycogen metabolism and cell cycle progression.
Biosynthesis	YLR260w	LCB5, L8479.7	Sphingoid long chain base (LCB) kinase (involved in sphingolipid biosynthesis).
	YNR041c	COQ2, N3419	Para-hydroxybenzoate:polyprenyl transferase
	YOL143c	RIB4, A0E169, O0467	6,7-dimethyl-8-ribityllumazine synthase (DMRL synthase) (part of the riboflavin biosynthetic pathway)
Other functions	YDR138w	HPR1, TRF1, YD9302.14	Hyperrecombination protein related to Top1p, not essential for repair or meiosis
	YGL107c	G3075	Strong similarity to hypothetical protein YBR238c
	YJL188c	J0403	Hypothetical ORF
	YNL225c	CNM67, CNM1, N1264	Cytoskeletal structural protein, deletion causes chaotic nuclear migration
	YPL045w	VPS16, VPT16, VAM9, (COS5), SVL6, P7102.06	Vacuolar sorting protein
	YPL059w	GRX5, LPE13	Member of the subfamily of yeast glutaredoxins (Grx3, Grx4, and Grx5).

Table 2-4 (Part 5): Genes identified as being essential for stationary phase (information collected from the MIPS and SGD databases.

Genes YDL068c and YGR160w have no assigned function and are not similar to any known gene. These ORFs overlap other known genes and have been identified as ORFs that are probably not genes in a reannotation of the *S. cerevisiae* genome. However the genes with which these ORFs overlap (YDL069c/*CBS1* and YGR159c/*NSR1* respectively) were not identified as genes that are essential for stationary phase. This suggests that YDL068c and YGR160w are in fact real ORFs. If these two genes were

hypothetical genes and the overlapping genes are the genes with the essential stationary phase function then the overlapping genes should have also been identified in this study.

The proper processing of certain mRNAs seems to be important for stationary phase survival. YDL107w, YHL038c, and YLR203c are all involved in the splicing and maturation of *COB* and *COX* gene mRNAs, and YMR064w is involved in ATPase mRNA expression. While, more generally, YDL044c and YIR021w are involved in the stability/maturation of mitochondrial RNAs. Two genes, YLR067c and YDR197w, are involved in translation prior to the actual start of translation itself. YLR067c is required for stability and translation of *COX1* mRNA, and YDR197w is a translational activator protein for cytochrome B.

Many ribosomal proteins were identified as being essential for the maintenance of viability in stationary phase, with the proteins from the mitochondrial ribosome being notably prominent (Tables 2-a and 2-3). While ORF YNR036c has yet had no function assigned to it, it has a strong similarity to ribosomal protein S12 so may be a ribosomal protein. The deletion of several other genes that code for transcriptionally associated proteins such as tRNA synthetases may also affect the ability of the ribosome to produce proteins (table 2-4)..

A large number of mitochondrial ribosomal genes were identified in the screen but not every known gene of this classification was included. Of these, seven were originally classified as not having a stationary phase defect due to the absence of growth in one or more of the cultures, one because of contamination of one of the cultures, and the rest because the two haploid mating types tested gave opposing results (i.e. one remained viable and the other had lost viability after 3 months).

Ribosome		Mitochondrial Ribosome	
		YCR046c, YDL045w-a, YPR166c	
Small Subunit	**Large Subunit**	**Small Subunit**	**Large Subunit**
	YGL135w	YDR175c, YDR337w, YGL129c, YJR113c, YMR158w and YNR037c.	YBR268w, YCR003w, YDL202w, YDR115w, YDR237w, YDR405w, YGR076c, YGR220c, YHR147c, YJL063c, YKL138c, YKR006c, YKR085c, YMR193w, YNL177c, YPL173w, and YPR100w.

Table 2-5: Ribosomal genes identified as causing a loss in viability in stationary phase when knocked out.

Twenty-six gene knockout strains were retested (results appendix 2, list of genes appendix 3). Nineteen of the strains were still viable after 93 days in stationary phase. No conclusion could be drawn about six of the strains while the YNL185c gene knockout strain indicated a possible stationary phase defect (Table 2-6). This conclusion is however not certain due to the loss of results for the two haploid strains of this gene knockout.

Stationary phase essential ORFs - *ORF Pairs*

When the list of the ORFs required for stationary phase is examined several instances of adjacent ORFs both giving a stationary phase defect when knocked out can be found. It is possible that in these cases both of the ORFs are required for stationary phase. However it is also possible that the deletion of one of the ORFs is somehow impacting upon the function of the second for example by deleting an essential part of the promoter of the second ORF.

Repeat	Growth of culture after 7 days				Growth of sample after 93 days			
	MATa	MATα	Hetero.	Homo.	MATa	MATα	Hetero.	Homo.
1	No	Yes	Yes	Yes	No result	No result	Confluent	18
2	No	Yes	Yes	Yes	No result	No result	Confluent	37

Table 2-6: The results of retesting (in YPD medium) the YNL185c ORF knockout in the MATa and MATα haploid and the heterozygous (hetero.) and homozygous (homo.) diploid strains.

YDL044c and YDL045w-a (Diagram 2-5) are respectively a mitochondrial protein involved in mRNA splicing and protein synthesis and a mitochondrial ribosomal protein. While the systematic names might suggest that these two ORFs are located next to each other, they are in fact separated by another ORF. Therefore the deletion of one of either YDL044c or YDL045w-a would not affect the other ORF.

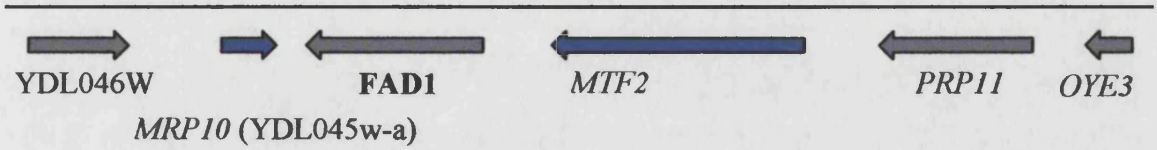


Diagram 2-5: A 6Kb section of Chromosome IV, showing YDL044c (MTF2) and YDL0045w-a (MRP10) in blue. Also shown are the PRP11, FAD1 YDL046w, and the 3'-termini of the SIR2 ORFs, which are adjacent to and between these two ORFs.

YDL067c is a cytochrome c oxidase subunit and YDL068w a questionable ORF. As YDL068w and YDL067c are both downstream of each other (Diagram 2-6), deletion of one ORF will not affect the transcription of the other. However a reannotation of *S. cerevisiae* genome (Wood, Rutherford et al. 2001) has suggested that YDL068w is not a real ORF as it overlaps the gene YDL069c (*CBS1*; Translational activator of cob mRNA). When tested the two haploid strains for YDL069c gave conflicting results. The MATa strain had no viability after 93 days, while the MATα strain showed viability. Results that gave no obvious result, such as this result, were not classified as giving a possible stationary phase defect and were not retested to confirm the defect. As YDL068w was classified as having a stationary phase defect (both haploid strains lose viability after 93 days), and is believed to not actually be a real gene, it would suggest that the result for YDL069c is a false negative and therefore that YDL069c is essential for stationary phase.

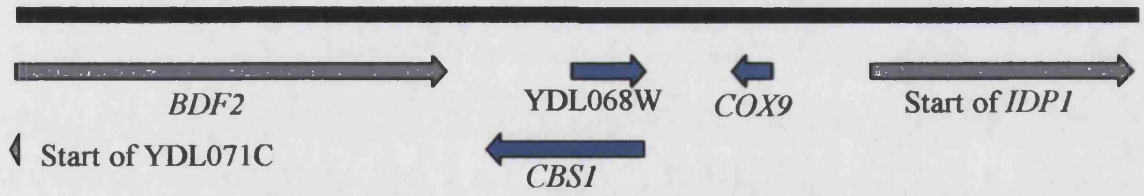


Diagram 2-6: A 5Kb section of Chromosome IV, showingYDL067c (*COX9*), YDL068w and YDL069c (*CBS1*) in blue. Also shown are the BDF2, and the 5'-termini of the IDP1 and YDL071c ORFs, which are adjacent to these overlapping and adjacent ORFs.

YNR036c and YNR037c are a hypothetical ORF and a gene that codes for a mitochondrial ribosome protein of the small subunit respectively. As YNR036c is downstream of YNR037c (Diagram 2-7), a deletion of YNR037c might delete part of the promoter for YNR036c resulting in a false conclusion that YNR037c is essential for

stationary phase. However, given the large number of mitochondrial ribosomal proteins identified by this investigation, it is probable that the ORF YNR037c (a mitochondrial ribosomal gene) is required for stationary phase which leads to the conclusion that both YNR036c and YNR037c code for proteins that are required for stationary phase.

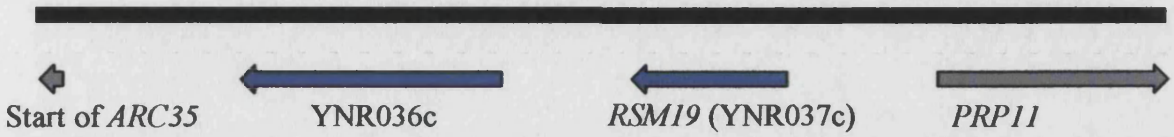


Diagram 2-7: A 2Kb section of chromosome XIV, showing the ORFs YNR036c and YNR037c (*RSM19*) in blue. Also shown are the 5' termini of the ORFs *DBP6* and *ARC35* which are adjacent to this pair of genes.

YPL172c and YPL173w (Diagram 2-8) are a putative farnesyl transferase and a mitochondrial ribosomal protein. Both these genes are downstream of each other. Therefore the deletion of one ORF will not affect the translation of the other. The putative farnesyl transferase is required for heme A synthesis, which is in turn required for an essential posttranslational stage in assembly of cytochrome oxidase, which links this gene to respiration (and thus possibly essential for stationary phase viability). Since one gene is functionally linked to respiration and the other to a mitochondrial ribosomal protein (of which a large number seem to be essential for stationary phase) it is likely that both of these genes are essential for stationary phase.

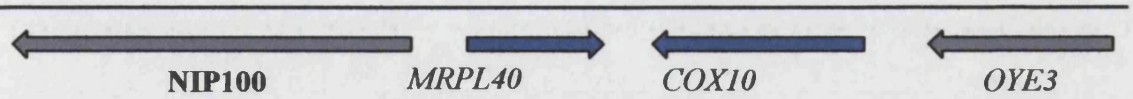


Diagram 2-8: A 7.5Kb section of chromosome XVI, showing the ORFs YPL172c (*COX10*) and YPL173w (*MRPL40*) in blue. Also shown are the ORFs *NIP100* and *OYE3* that are adjacent to this pair of genes.

YPR099c and YPR100w are both hypothetical ORFs (Diagram 2-9). In a reannotation (Wood, Rutherford et al. 2001) of the *S. cerevisiae* genome it was concluded that YPR099c was a spurious ORF and that YPR100w was the actual ORF in this overlapping pair. Therefore a deletion of YPR099c would also delete part of a stationary phase essential ORF (YPR100w) and would therefore appear to be essential as well.

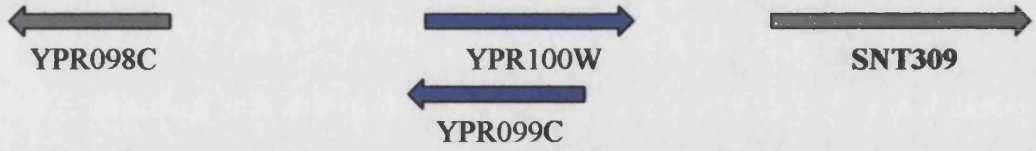


Diagram 2-9: A 2.4Kb section of chromosome XVI, showing the ORFs YPR099c and YPR100w in blue. Also shown are the ORFs *SNT309* and YPR098c that are adjacent to this pair of genes.

Stationary phase essential ORFs - *Genomic distribution*

Of all the gene knockouts tested, 99 can be positively identified as having a defect in stationary phase when the gene is knocked out. Based on the number of genes that were identified as having a stationary phase defect (Table 2-7) the expected number of genes per

chromosome with a stationary phase defect can be calculated. This calculation uses the assumption that genes with a stationary phase gene defect occur at random throughout the genome, and are not linked to other stationary phase essential genes or a particular chromosome.

Chromosome	Number of genes:		Total number of genes tested
	with a defect	without a defect	
A	0	89	89
B	2	413	415
C	2	158	160
D	23	759	782
E	2	248	250
F	1	123	124
G	12	529	541
H	4	226	230
I	1	200	201
J	6	342	348
K	5	331	336
L	7	505	512
M	12	459	471
N	7	400	407
O	4	540	544
P	17	456	473
Total	105	5778	5883

Table 2-7: The number of genes per chromosome with and without a stationary phase defect, tested for a stationary phase defect.

The expected number of genes on a chromosome with a stationary phase defect is the fraction of the total number of genes in the genome that are found on the chromosome in question, multiplied by the total number of genes with a stationary phase defect:

Expected number of genes with a stationary phase defect on a chromosome

$$= \left(\frac{\text{Number of genes on the chromosome}}{\text{Total number of genes}}\right) \text{ X Total number of genes with a stationary phase defect}$$

The χ^2-test can be used to determine whether the number of genes with a stationary phase defect seen on each chromosome is significantly different from the numbers predicted. As the test will be comparing whether the observed number of genes with and without a stationary phase defect per chromosome are similar to the expected numbers there will be only one degree of freedom for the χ^2 test. At one degree of freedom the χ^2 test with Yates' correction should be used:

$$\chi^2_\alpha = \Sigma\left(\frac{(|\text{Observed} - \text{Expected}| - 0.5)^2}{\text{Expected}}\right)$$

F this calculation the observed/expected numbers and the null hypothesis are:

Observed	the number of genes on a chromosome with a stationary phase defect.
Expected	the expected number of genes with a stationary phase defect on a chromosome.
Null Hypothesis (H_0)	"*The genes with a stationary phase defect occur at random throughout the* Saccharomyces cerevisiae *genome, and there is no bias towards genes on any one particular chromosome*"

Using the calculated value of χ^2_α the probability that the difference between observed and expected numbers is occurring by random, $P(\alpha)$, can be calculated. For the occurrence of genes with a stationary phase defect when the gene is knocked out on each chromosome this calculation can be seen in Table 2-8. If $P(\alpha) < 0.05$ it can be concluded that the observed and expected numbers are significantly different at the 95% level. As there are two values for each chromosome (*with defect* and *without defect*) the degrees of freedom with which $P(\alpha)$ is calculated will be 1.

Chrom.	Observed (O)		Expected (E)		$((\|O-E\|-0.5)^2)/E$			P(α)
	w/	w/o	w/	w/o	w/	w/o	total	
A	0	89	1.59	87.41	0.75	0.01	0.76	0.38
B	2	413	7.41	407.59	3.25	0.06	3.31	0.07
C	2	158	2.86	157.14	0.04	0.00	0.05	0.83
D	23	759	13.96	768.04	5.23	0.10	5.32	0.02
E	2	248	4.46	245.54	0.86	0.02	0.88	0.35
F	1	123	2.21	121.79	0.23	0.00	0.23	0.63
G	12	529	9.66	531.34	0.35	0.01	0.36	0.55
H	4	226	4.11	225.89	0.04	0.00	0.04	0.84
I	1	200	3.59	197.41	1.21	0.02	1.24	0.27
J	6	342	6.21	341.79	0.01	0.00	0.01	0.91
K	5	331	6.00	330.00	0.04	0.00	0.04	0.84
L	7	505	9.14	502.86	0.29	0.01	0.30	0.58
M	12	459	8.41	462.59	1.14	0.02	1.16	0.28
N	7	400	7.26	399.74	0.01	0.00	0.01	0.93
O	4	540	9.71	534.29	2.79	0.05	2.85	0.09
P	17	456	8.44	464.56	7.69	0.14	7.83	0.01
Total	105	5778	105	5778				

Table 2-8: Comparison of the observed (O) and Expected (E) number of genes with (w/) and without (w/o) a stationary phase defect on each chromosome when the gene is knocked out using the χ^2 test (P(α) was calculated using degrees of freedom = 1).

For the majority of chromosomes the expected number of genes with and without a stationary phase defect when knocked out was observed. For chromosomes D and P this was not the case. Gene knockouts with stationary phase defects from chromosomes D and

P were over-represented. A possible explanation for this is that the ORFs are not randomly distributed. There is evidence that in *S. cerevisiae* co-regulated genes are linked (Hurst, Williams et al. 2002; Pál and Hurst 2003). As the set of genes appears to be biased towards mitochondrial and ribosomal functions, a relatively large proportion of the genes could be expected to be expressed in a very similar manner. This therefore could cause an over-representation of one or more areas of the genome if co-ordinately regulated genes grouped together.

There are three functional classifications of genes that appear to be over-represented in this set of genes (tables 2-8 and 2-9). If the occurrence of genes in these classifications on the chromosomes D and P is examined using a χ^2 test it can be seen that none of these classifications are over represented on either chromosome (Table 2-9). Which might suggest that the over-representation of genes from the chromosomes D and P is not due to the clustering of co-ordinately regulated genes as genes of a similar function can be expected to be regulated in a similar manner to each other.

Classification	**Observed**		**Expected**		**χ^2 test; P(α)**	
	Chrom. D	**Chrom. P**	**Chrom. D**	**Chrom. P**	**Chrom. D**	**Chrom. P**
Protein Synthesis	49	33	46	28	0.733	0.381
Energy	25	22	32	19	0.265	0.585
Subcellular localization	301	169	289	175	0.415	0.594

Table 2-9: Occurrence of genes from three functional classifications on the chromosomes D and P.

Stationary phase essential ORFs - *Functional classification*

All the genes in the *S. cerevisiae* genome have been classified according to what is known about the function of the protein for which the gene encodes. These lists of genes are available on the MIPS website: http://mips.gsf.de/proj/yeast/CYGD/db/index.html, and the different categories can be seen in column 1 of table 2-10. It should be noted that genes could occur in more than one of the categories. Using the total number of genes in the genome the percentage of genes occurring in each category can be calculated:

Percentage of the total number of genes in each category

$$= \left(\frac{\text{Number of genes in the category}}{\text{Total number of genes in the genome}}\right) \times 100$$

These figures can then be used to calculate the number of the 124 genes with a stationary phase defect that would be expected to appear in each category:

Expected number of gene knockouts with a stationary phase defect in each category

$$= \text{Number of knockouts with a stationary phase defect} \times \left(\frac{\text{Percentage of genes in the category}}{100}\right)$$

The expected number of genes and the observed number of genes occurring in each category which have a stationary phase defect when knocked out are shown in Table 2-10.

The χ^2-test can be used to determine whether the number of genes with a stationary phase defect seen in each category is significantly different from the numbers predicted. As with testing for the distribution of the genes across the chromosomes the χ^2-test with Yates' correction is used. For the genes with stationary phase defects:

$$\chi^2_\alpha = \Sigma\left(\frac{(|\text{Observed} - \text{Expected}| - 0.5)^2}{\text{Expected}}\right)$$

Observed The number of genes in each category with a stationary phase defect.

Expected The expected number of genes with a stationary phase defect on a chromosome.

Null Hypothesis (H_0) "*The number genes with a stationary phase defect occurring in each functional classification relative to the total number of gene knockouts is the same as the total number of genes in each category relative to the total number of genes in the* Saccharomyces cerevisiae *genome*"

Using the calculated value of χ^2_α the probability that the difference between observed and expected numbers is occurring by random, $P(\alpha)$, can be calculated. For the occurrence of genes with a stationary phase defect in each functional classification this calculation can be seen in Table 2-10. If the $P(\alpha) < 0.05$ it can be concluded that the observed and expected numbers are significantly different at the 95% level. As there are two values for each chromosome (*with defect* and *without defect*) the degrees of freedom with which $P(\alpha)$ is calculated will be 1.

For the majority of functional classifications the numbers of genes with a stationary phase defect when knocked out is not significantly different. However for four of the classifications there is a significant difference at the 95% level. The three functional classifications "Subcellular Localisation", "Energy" and "Protein Synthesis" are over-represented while the classification "Unclassified Proteins" is under-represented.

The functional categories can be broken down into different subcategories, and each subclassification can be tested for over- and under-representation using χ^2 (Table 2-11). The unclassified proteins functional classification cannot be broken down into subcategories, as the function of these ORFs are not yet known. However it is surprising that this set of ORFs is underrepresented in this study. This might suggest that the processes that are required for stationary phase are not mediated through any novel pathways but use pathways that have already been discovered.

Functional Classification	total in list	Expected		Observed		P(α)
		with defect	without defect	With defect	without defect	
CELL CYCLE AND DNA PROCESSING	628	10.41	617.59	12	616	0.73
CELL FATE	428	7.09	420.91	4	424	0.33
CELL RESCUE, DEFENSE AND VIRULENCE	279	4.62	274.38	1	278	0.14
CELLULAR COMMUNICATION/SIGNAL TRANSDUCTION	60	0.99	59.01	0	60	0.62
CELLULAR TRANSPORT AND TRANSPORT MECHANISMS	496	8.22	487.78	4	492	0.19
CLASSIFICATION NOT YET CLEAR-CUT	116	1.92	114.08	1	115	0.76
CONTROL OF CELLULAR ORGANIZATION	210	3.48	206.52	2	208	0.6
ENERGY	253	4.19	248.81	16	237	0
METABOLISM	1067	17.69	1049.31	13	1054	0.32
PROTEIN ACTIVITY REGULATION	14	0.23	13.77	0	14	0.57

Table 2-10 (part 1): Comparison of the observed (O) and Expected (E) number of genes with a stationary phase defect when the gene is knocked out in each functional classification (from the MIPS database; http://mips.gsf.de) using the χ^2 test. Expected numbers of genes and calculated values of P(α) given to 2 d.p.

Functional Classification	total in list	Expected		Observed		P(α)
		with defect	without defect	With defect	without defect	
PROTEIN FATE (folding, modification, destination)	596	9.88	586.12	10	586	0.9
PROTEIN SYNTHESIS	360	5.97	354.03	42	318	0
PROTEIN WITH BINDING FUNCTION OR COFACTOR REQUIREMENT (structural or catalytic)	5	0.08	4.92	0	5	0.14
REGULATION OF / INTERACTION WITH CELLULAR	200	3.31	196.69	4	196	0.92
SUBCELLULAR LOCALISATION	2259	37.44	2221.56	69	2190	0
TRANSCRIPTION	772	12.8	759.2	15	757	0.63
TRANSPORT FACILITATION	314	5.2	308.8	3	311	0.45
TRANSPOSABLE ELEMENTS, VIRAL AND PLASMID	117	1.94	115.06	0	117	0.3
UNCLASSIFIED PROTEINS	2400	39.78	2360.22	18	2382	0
Total	10574	175.26	10398.74	214	10360	0

Table 2-10 (part 2): Comparison of the observed (O) and Expected (E) number of genes with a stationary phase defect when the gene is knocked out in each functional classification (from the MIPS database; http://mips.gsf.de) using the χ^2 test. Expected numbers of genes and calculated values of P(α) given to 2 d.p.

FUNCTIONAL		Observed		Expected		((|O-E|-0.5)^2)/E			P(α)
Subcategory/ sub- subcategory		in	NOT in	in	NOT in	in	NOT in	Total	
UNCLASSIFIED PROTEINS		18	87	39.75	65.25	11.36	6.92	18.27	0
SUBCELLULAR		69	36	37.43	67.57	25.8	14.29	40.09	0
cell wall		0	105	0.63	104.37	0.03	0	0.03	0.87
plasma membrane		0	105	2.4	102.6	1.51	0.04	1.54	0.21
cytoplasm		2	103	9.18	95.82	4.86	0.47	5.33	0.02
cytoskeleton		0	105	1.81	103.19	0.95	0.02	0.96	0.33
centrosome		1	104	0.51	104.49	0	0	0	0.98
endoplasmic reticulum		1	104	2.6	102.4	0.47	0.01	0.48	0.49
Golgi		0	105	1.36	103.64	0.54	0.01	0.55	0.46
intracellular transport vesicles		0	105	0.7	104.3	0.06	0	0.06	0.81
nucleus	(all)	5	100	12.83	92.17	4.19	0.58	4.77	0.03
	chromosome	0	105	0.73	104.27	0.07	0	0.07	0.79
mitochondrion		60	45	6.07	98.93	470.66	28.86	499.52	0

Table 2-11 (part 1): Comparison of the observed (O) and Expected (E) number of genes with a stationary phase defect when the gene is knocked out in each subclassification of the significantly over- and underrepresented functional classification (from the MIPS database; http://mips.gsf.de) using the χ^2 test. Expected numbers of genes and calculated values of P(α) given to 2 d.p.

FUNCTIONAL CLASSIFICATION		Observed		Expected		((|O-E|-0.5)^2)/E			P(α)
Subcategory/ sub- subcategory		in	NOT	in	NOT	in	NOT	Total	
peroxisome		0	105	0.65	104.35	0.03	0	0.03	0.86
endosome		0	105	0.22	104.78	0.38	0	0.38	0.54
vacuole or lysosome		1	104	0.98	104.02	0.23	0	0.24	0.63
extracellular / secretion proteins		0	105	0.33	104.67	0.09	0	0.09	0.77
prokaryotic cell membrane (inner membrane of gram -ve bacteria)		0	105	0.02	104.98	14.1	0	14.1	0
other subcellular localisation		0	105	0.13	104.87	1.02	0	1.02	0.31
ENERGY		16	89	4.18	100.82	30.7	1.27	31.97	0
glycolysis and gluconeogenesis		0	105	0.58	104.42	0.01	0	0.01	0.92
pentose-phosphate pathway		0	105	0.15	104.85	0.83	0	0.83	0.36
tricarboxylic-acid pathway		4	101	0.41	104.59	22.98	0.09	23.07	0
electron transport and membrane-associated energy conservation	(all)	0	105	0.03	104.97	6.57	0	6.58	0.01
	accessory proteins of electron transport and membrane-associated energy conservation	0	105	0.02	104.98	14.1	0	14.1	0
	other electron transport and membrane-associated energy conservation proteins	0	105	0.02	104.98	14.1	0	14.1	0

Table 2-11 (part 2): Comparison of the observed (O) and Expected (E) number of genes with a stationary phase defect when the gene is knocked out in each subclassification of the significantly over- and underrepresented functional classification (from the MIPS database; http://mips.gsf.de) using the χ^2 test. Expected numbers of genes and calculated values of P(α) given to 2 d.p.

FUNCTIONAL		Observed		Expected		((|O-E|-0.5)^2)/E			P(α)
Subcategory/ sub-		in	NOT in	in	NOT in	in	NOT in	Total	
respiration		11	94	1.46	103.54	56.05	0.79	56.84	0
fermentation		0	105	0.55	104.45	0	0	0	0.95
metabolism of energy reserves		1	104	0.61	104.39	0.02	0	0.02	0.88
glyoxylate cycle		0	105	0.1	104.9	1.61	0	1.61	0.2
oxidation of fatty acids		0	105	0.12	104.88	1.27	0	1.27	0.26
other energy generation		1	104	0.27	104.73	0.21	0	0.21	0.65
PROTEIN SYNTHESIS		42	63	5.95	99.05	212.39	12.76	225.15	0
ribosome biogenesis		25	80	3.56	101.44	123.01	4.32	127.33	0
Translation	(all)	3	102	1.06	103.94	1.95	0.02	1.97	0.16
	initiation	0	105	0.03	104.97	6.57	0	6.58	0.01
	elongation	1	104	0.03	104.97	6.57	0	6.58	0.01
translational control		7	98	0.51	104.49	69.74	0.34	70.09	0
aminoacyl-tRNA-synthetases		5	100	0.61	104.39	24.63	0.14	24.78	0
other protein-synthesis		2	103	0.27	104.73	5.75	0.01	5.76	0.02

Table 2-11 (part 3): Comparison of the observed (O) and Expected (E) number of genes with a stationary phase defect when the gene is knocked out in each subclassification of the significantly over- and underrepresented functional classification (from the MIPS database; http://mips.gsf.de) using the χ^2 test. Expected numbers of genes and calculated values of P(α) given to 2 d.p.

The subcellular localisation classification can be spilt into 16 different subcategories. Of these subcategories the "Cytoplasm", "Nucleus", "Mitochondrion" and "Prokaryote Cell Membrane" are all significantly under- or over-represented. Both cytoplasmic and nuclear located proteins are underrepresented. The nucleus subcategory is further subcategorised but the numbers of ORFs appearing in the chromosome subcategory are not significantly different from the expected values. Mitochondrially located proteins are significantly over represented. Approximately 10 times the expected number of ORFs was observed in this category. While the under-representation of the prokaryote cell membrane subcategory is a significant result, it can be ignored. While no ORFs from this category were found to have a stationary phase defect when knocked out, only 0.02 ORFs were expected to occur in this category. Rounding the expected number of ORFs off to the nearest integer results in 0 ORFs (which is the number of ORFs observed).

The Energy functional classification consists of 10 subcategories. Of these 3 have significantly different numbers of ORFs than what was expected. ORFs from the TCA cycle are represented at approximately 10-fold greater than what is expected. ORFs from the respiration subcategory are over represented at approximately the same scale as those from the TCA cycle subcategory. The electron transport and membrane-associated energy conservation subcategory is under represented. However while no ORFs from this category were found to be essential for stationary phase, approximately no ORFs were expected to be classified in this category. Like the prokaryote cell membrane category this result is a misleading result and can be ignored.

Four of the five subcategories in the protein synthesis functional classification are significantly over- or under-represented. Ribosome biogenesis, translational control, amino-tRNA-synthetases and other protein synthesis activities are all over-represented by 5- to 10-fold from what would be expected. While representation of the translation subcategory is not significantly different, the two sub-subcategories produce values of P(α)

that would suggest an over- or under-representation. However examination of the observed and expected numbers reveals that these results can be ignored as anomalous results.

Respiratory Competence and Stationary Phase

The results suggest that some aspect of mitochondrial function and/or respiration is playing an important part in stationary phase viability. To test this conclusion, mitochondrial mutants were generated by exposing *S. cerevisiae* FY1679 to ethidium bromide (Mahler and Perlman 1972). A total of 51 independently isolated *petite* mutants was obtained after ethidium bromide treatment, and were confirmed to be respiratory deficient by their failure to grow on YPG medium. When put through the same test for stationary phase as the ORF knockouts all 51 of the petite mutants suffered a complete loss of viability. This provides independent evidence of the importance of the mitochondria in stationary phase viability.

The citric acid cycle and oxidative phosphorylation along the respiratory chain both occur in the mitochondria. These pathways generate energy by oxidising reduced carbon sources. An obvious conclusion is that respiration is required for maintenance of stationary phase viability. There is evidence to support this hypothesis: Much of the energy production in stationary phase comes from mitochondrial respiration; and increasing oxidative damage defences (as most cellular ROS come are generated in the mitochondria) or pre-adapting a cell to respiratory growth increases lifespan (Werner-Washburne, Braun et al. 1993; Longo, Gralla et al. 1996; Longo, Liou et al. 1999; MacLean, Harris et al. 2001)

Contrary to this hypothesis, the screening for stationary phase essential ORFs conducted in this investigation seems to indicate that respiratory function is not essential for the maintenance of viability in stationary phase. Some ORF knockouts which are known to be respiratory deficient do not show a loss of stationary phase viability (table 2-12).

	Stationary phase defect	No stationary phase defect
Respiratory proficient	Knockout strains of the ORFs: YBR179c, YDR178w, YHL038c, and YML129c	*S. cerevisiae* FY1679 (control); Knockout strains of the ORFs: YDR059c, YER131w, YKL169c, and YNL037
Respiratory deficient	*S. cerevisiae* FY1679 *petite* mutants; Knockout strains of the ORFs: YDR148c, YGR062c, YGR220c, and YMR228w	Knockout strains of the ORFs: YBR003w, YER017c, YKL087c, YOL071w, and YPL104w

Table 2-12: respiratory deficient and respiratory deficient strains can exist with stationary phase defective and stationary phase unaffected phenotypes (when grown in YPD medium).

To confirm this the MIPS database was consulted and ORF knockouts that are known to be respiratory deficient were retested for a defect in maintaining stationary phase viability. The results of this retest of 324 respiratory deficient knockout strains can be seen in appendix 2 and a summary of the results is shown in table 2-13. The results from this test of respiratory deficient strains show that many of the respiratory deficient strains (133 out of 324 tested strains) did not die during long-term starvation. This confirms that respiration is not essential for the maintenance of viability in stationary phase.

Knockout strain has a stationary phase defect	Retest	Initial screen
Yes	95	74
No	160	250
Results inconclusive	69	-
Total	324	324

Table 2-13: Summary of the retest of respiratory deficient ORF knockout strains.

Of the 324 ORF knockouts tested about 101 strains gave a different result to that observed during the screen of the entire SGDP set of knockouts (Initial screen). Of these contrary conclusions, 69 were due to inconclusive results during the retest (for example the two haploid strains gave conflicting results, strains were lost due to contamination, etc.). Almost all of the rest of the differences (31 ORF knockouts) were of ORFs that were originally classified as having no defect in the initial screen, but showed a definite or probable stationary phase defect in the retest. Results of the first screen of the SGDP set for these strains were re-examined and it was seen that twelve ORF knockouts were discarded as being stationary phase essential ORFs, ten knockouts gave mixed results, and one was lost due to contamination. Conclusions for a further eight strains could not be obtained due to one or both of the strains not growing. The remaining 10 strains appeared during the first screen of the SGDP set of ORF knockouts not to have had a stationary phase defect.

Only one ORF knockout (YGR220c) that was originally determined to have a stationary phase defect was found not to give a defect during the retest of respiratory deficient knockout strains. This ORF knockout when retested in YPD and other media was found to have a stationary phase defect. Therefore the growth that was observed for this strain was probably due to contamination and the non-defect result in the test of respiratory

deficient knockout strains is an anomalous result and can be ignored. Despite some inconclusive results the overall conclusion is clear – over one third of respiratory deficient mutants do not show a significant loss of viability in stationary phase.

Another piece of evidence that suggests that respiration is not essential for stationary phase is the effect of adding reduced glutathione to the stationary phase cultures. Glutathione did not rescue the stationary phase defect of most of the knockout strains, indicating that for these strains oxidative damage is not the factor causing the loss of viability during stationary phase. During stationary phase cells are believed to generate energy from respiration (Lillie and Pringle 1980; Longo, Gralla et al. 1996; Roy and Ghosh 1998; Silljé, Paalman et al. 1999; Samokhvalov, Ignatov et al. 2004). Respiration is known to be the largest source of reactive oxygen species in the cell, and cells with less effective or no protection against oxidative damage cannot survive long-term stationary phase (Longo, Liou et al. 1999; Jakubowski, Bilinski et al. 2000). The inability of a oxidative damage protectant to reverse the stationary phase defects of knockout strains suggests that the loss of viability is not due to oxidative damage, and this suggests that there is little or no respiration occurring.

Discussion

Different media

The results obtained from the different media vary considerably (table 2-3). There are no genes that, when knocked out, cause a stationary phase defect in all of the media used. This would suggest that none of the ORF knockouts disrupt a process that is essential for the maintenance of viability during stationary phase. If this is the case and there is no single process that is essential, the environment in which the cells enter/maintain stationary phase becomes an important factor. Individual ORF knockouts cause a lethal phenotype only as a secondary effect.

Rescue of strains in low glucose medium

Why do ORF knockout strains survive stationary phase better when grown in low glucose medium? The rescue of the knockouts by growth in YP(0.3D) could be due to an increase in respiration. Low glucose conditions could be increasing respiration via Sir2 (Lin, Kaeberlein et al. 2002) but this is unlikely because not all the knockout strains identified were able to respire, a fact that was confirmed by retesting of known respiratory deficient knockout strains.

Could osmolarity of the medium be important? A dextrose solution of 0.3% will have an osmolarity of 0.0016 osmolar compared to 0.0111 osmolar for a 2% dextrose solution. This reduction of osmolarity however will only be present under non-stationary phase conditions, because whatever the starting concentration of dextrose there will be no dextrose in either medium when the cells enter stationary phase. Furthermore the strains were tested in YPD medium with 1 M sorbitol added to act as a osmotic protectant. While some of the strains were rescued by the addition of the sorbitol, a similar number were not rescued. This result could be interpreted as an indication that osmotic stress can be a factor in surviving stationary phase but it is not essential.

Difference between YPD and SC media

The results obtained from growing ORF knockout strains in YPD and SC media are generally the same. However some ORF knockout strains able to survive stationary phase in spent YPD medium were found to be unable to survive stationary phase when grown in SC medium. This difference between the media was even more pronounced when the low glucose version of each of the media was used. Strains rescued by growth in YP(0.3D) medium were not rescued when incubated through long-term stationary phase in SC(0.3D) medium. This is possibly because the cells must divert energy/material away from preventing/repairing cellular damage to producing substances that could be imported from the medium when grown in YPD medium. This would therefore reduce the ability of a cell

to maintain viability in stationary phase as its stores of energy and materials would be consumed at an increased rate. However, attempting to explain increased loss of viability by a lower level of nutrients in the medium does not account for 13 strains that cannot survive stationary phase when grown in YPD but can when grown in the less rich SC medium. The 13 genes involved do not appear to be related to each other by function which would seem to indicate that the effect observed is not due to a common cellular process that these genes are involved with.

ORFs of known and unknown function

A more detailed discussion of genes identified as being essential for stationary phase is discussed in chapter 6.

Chapter 3

Stationary phase phenotypes

Summary

The range of phenotypic properties of knockouts of ORFs essential for stationary phase survival were investigated. A visual examination showed that all the ORF knockouts arrested as unbudded cells on starvation. However, most of the ORFs did not develop an increased heat shock resistance, characteristics of normal stationary phase entry, and a subset of these ORFs did not accumulate glycogen and trehalose, another normal characteristic upon entry to stationary phase. These results indicate that the knockout strains can recognise but not properly respond to starvation conditions. Incubation of stationary phase cultures in water has been shown to reduce the loss of cell viability. When this effect was studied it was shown only to work for short-term rather than long-term incubation.

Phenotypes of Stationary Phase Essential Genes

Various phenotypes are associated with cells that are entering or have entered stationary phase. Amongst other changes are their morphology, increased levels of resistance to heat shock, and increased levels of the carbohydrates glycogen and trehalose (Werner-Washburne, Braun et al. 1993; Werner-Washburne, Braun et al. 1996; Herman 2002). As well as being induced by entry into stationary phase, the heat shock response can affect the stationary phase viability of a cell. In some cases over-activation of the heat shock response can increase the length of time that a cell can survive stationary phase (Harris, MacLean et al. 2001). The more heat shock resistant a culture is the better the cells

in a culture will be able to survive a short incubation under high temperature conditions (Hampsey 1997).

The kinetics of accumulation of glycogen and trehalose differ. Glycogen levels start to rise before, and peak at, the diauxic shift when glucose becomes exhausted. The accumulation of trehalose does not begin until onset of the diauxic shift. Once in stationary phase the levels of both carbohydrates fall. This difference in accumulation/usage patterns suggests they have different functions. Glycogen is believed to be a reserve carbohydrate while trehalose is more important as a stress protectant (Werner-Washburne, Braun et al. 1993).

If a cell is unable to enter stationary phase properly some or all of these phenotypes associated with stationary phase will be absent. By examining the phenotypes being displayed by ORF knockout strains in stationary phase it was predicted that it might be possible to identify strains that did not seem to be responding properly to stationary phase conditions. Improper responses could provide a possible explanation for the loss of viability in stationary phase.

The incubation of cells in stationary phase in water instead of spent medium has been shown to prolong stationary phase survival (Granot and Snyder 1991; Granot and Snyder 1993; MacLean, Harris et al. 2001). This increased ability to survive stationary phase seems to be linked to respiration and carbon source that in turn may be linked to the ability of stationary phase essential gene null mutants to respire and be rescued by growth on low glucose medium.

Materials and Methods

Strains used

ORF knockout strains from the *Saccharomyces* Genome Deletion Project (SGDP) were used (Table 3-1) and *S. cerevisiae* FY1679 was used as a control.

Experiment	ORF knockout strains used
Heat shock resistance	105 ORF knockouts identified as having a stationary phase defect in this investigation.
Glycogen and trehalose levels	YDL044c, YDR194c, YJL188c, YJR144w, YKR006c, YKR024c, YML061c, YMR064w, YNL139c, and YNR036c.
Loss of cell viability	YDR194c, YDR405w, YNL177c, YCR003w, YDL202w, YGL135w, YGL220c, YJR113c, and YKR085c.

Table 3-1: SGDP ORF knockout strains used.

Heat shock resistance

Cultures of *S. cerevisiae* gene knockout strains were grown in 5 ml of YPD medium with shaking at 28°C in capped test tubes for 7 days. A 0.55 ml sample of each culture was taken at 7 days and used to start a second culture (5 ml YPD medium in a capped test tube) which was grown under the same conditions. A sample from one of the 7-day old cultures for each ORF knockout was also examined under a microscope and the budding index (the fraction of the total number of cells that are budding, multiplied by 100) determined. After 3 hours the budding index of one of the newly inoculated cultures for each ORF knockout (the same mating type/ploidy as was used to monitor the budding index of the 7-day old culture) was determined to confirm that the culture was in exponential phase. Using a *Hybaid OmniGene Temperature cycler,* 200 µl samples of the stationary (7-day) and exponential (growing) cultures were heat shocked at 55°C for 3 hours. After 0, 1, 2 and 3 hours of heat shock each of the samples were briefly vortexed to resuspend the cells and a 4 µl sample taken. Each of the 4 µl samples was spotted onto solid YPD medium. The growth of these samples was examined after 2 days growth at 28°C.

Cellular glycogen and trehalose levels

Glycogen and trehalose levels were tested based on the method of Parrou & François (Parrou and François 1997). Stationary phase cell cultures (1 week post-inoculation) were used, and a 200 µl sample of a 10^{-5} dilution was grown on solid YPD medium to give an estimate of cell number in the culture. Cell cultures (1.5 ml) were collected by centrifugation (2 min at 12,000 rpm), and were resuspended in 0.25 ml of 0.25 M Na_2CO_3. The cell suspensions were incubated at 95°C for 4 hours. Acetic Acid (0.15 ml, 1 M) and Sodium Acetate (0.6 ml, 0.2 M, pH 5.2) were then added to bring the mixture to pH 5.2.

Two overnight digestions of 140 µl samples were prepared form each of the mixtures. One sample was digested with *Aspergillus niger* amyloglucosidase (1.2 U/ml, at 57°C), the other with trehalase (0.05 U/ml at 37°C). For each mixture the level of glucose in each of the treated and an untreated sample was measured using a D-glucose testing kit from R-biopharm (*Hexokinase/Glucose-6-Phosphate Dehydrogenase*; Order number: 127 183) following the manufacturer's instructions.

Appearance of stationary phase cells

ORF knockout strains grown into stationary phase in YPD medium were shaken gently to resuspend the cell culture, then a sample was removed and examined using an *Olympus BH2* light microscope at x40 magnification.

Loss of cell viability

Each strain of *S. cerevisiae* was inoculated into 100 ml of YPD medium in a 500 ml conical flask with a foam bung. The cultures were grown at 28°C, with shaking, for one week. After the weeks growth in the conical flask each culture was split into two. The first half was split into thirteen 2.5 ml volumes and put into 4.5 ml Falcon tubes. The second half was washed and then resuspended in one volume of sterile distilled water. The second

half was then split into thirteen 2.5 ml volumes and each volume put into a 4.5 ml Falcon tube.

For both the unwashed (cell culture in spent medium) and washed (cell culture washed and resuspended in water) the Falcon tubes was stored static, and sealed shut at 28°C. Every week, from 1 week to 13 weeks post inoculation, the viability of each of the cultures (washed & unwashed) was determined. On the first week 10^{-4}, 10^{-5} and 10^{-6}-fold dilutions of each culture was plated out onto YPD medium and grown for two days at 28°C. The number of colonies growing was then used to estimate the number of viable cells in the culture. After the first week only one dilution, based on the results of the previous week, was used.

Gene knockout strains were used that had repeatably been shown to suffer a loss of viability in stationary phase over 13 weeks. Strains were chosen based on their ability or inability to respire and by whether the loss of viability in stationary phase phenotype could be rescued by growth in low glucose medium (YPD medium with 0.3% instead of 2% dextrose). The strains chosen are shown in Table 3-2. For each gene deletion four strains were used; two (*MAT***a** and *MAT*α) haploid strains, a heterozygous diploid and a homozygous diploid knockout strain. No strains were identified that lost viability in stationary phase, could respire, and were not rescued by growth in low glucose medium. The diploid strain FY1679 (which does not lose viability in stationary phase) was used as a control strain.

Gene knockout	Knockout strain can respire?	Growth in low glucose medium rescues loss of viability in stationary phase phenotype?
YDR194c	Yes	Yes
YDR405w	Yes	Yes
YNL177c	Yes	Yes
YCR003w	No	No
YDL202w	No	No
YGL135w	No	No
YGL220c	No	Yes
YJR113c	No	Yes
YKR085c	No	Yes

Table 3-2: *S. cerevisiae* gene knockout strains from the *Saccharomyces* Genome Deletion Project used in the investigation of loss of viability.

Results

Heat shock resistance of ORF knockouts with a stationary phase defect

S. cerevisiae FY1679 (Diagram 3-1) showed typical wild type characteristics. . The exponential phase cell culture lost viability in less than one hour of heat shock at 55°C. The stationary phase culture lost viability at 55°C but was still viable cells after 2 hours heat shock.

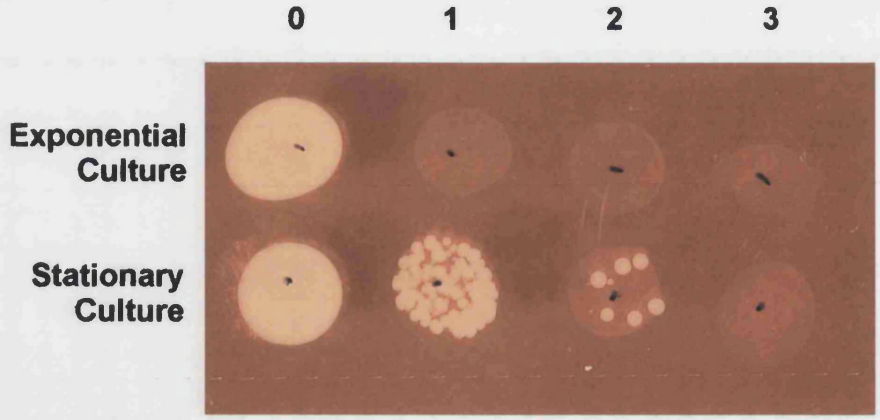


Diagram 3-1: Growth of 4 µl samples of a *Saccharomyces cerevisiae* FY1679 culture after 0, 1, 2, and 3 hours heat shock at 55°C.

Thus by comparing the growth of samples taken from exponential and stationary phase cultures after various times in heat shock it was possible to conclude in which phase the cell culture became more heat shock resistant. The resistance of the 105 ORF knockouts to heat shock at 55°C was tested. By comparing the four haploid and diploid types for each ORF knockout (Diagram 3-2) it was possible to make one of three conclusions:

1) The stationary phase culture is more heat shock resistant than exponential phase culture.
2) Neither culture is more heat shock resistant than the other.
3) The stationary phase culture is more heat shock resistant than exponential phase culture.

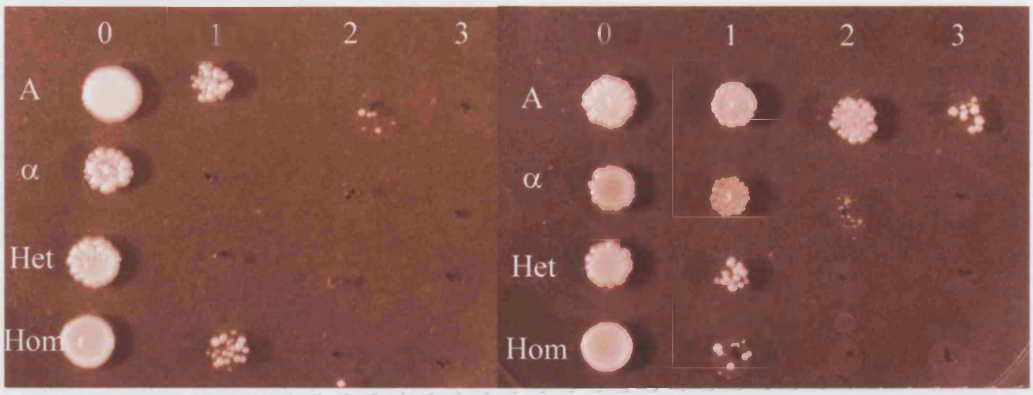


Diagram 3-2: Heat shock resistance of YDR300c knockout strains at 55°C from exponential (Left) and stationary (Right) phase cultures. Showing that YDR300c knockout strains are more heat shock resistant in stationary phase. Key: A = MAT**a** mating type, α = MATα mating type, Het = Heterozygous diploid, Hom = Homozygous diploid; 0/1/2/3 = Number of hours of heat shock.

The budding index of the cultures used was monitored to determine the phase of growth that the cultures were in. In exponential cultures the budding index will be high, while for cultures in stationary phase the budding index will be low. Theoretically when a culture is arrested the budding index should be 0% but in practice values of 10-20% are more usual (diagram 2-2).

For the knockout strains examined the stationary phase cultures had a average budding index of 13.5% ± 7.7 (± standard deviation, to 1 d.p.) (diagram 3-3). After 3 hours growth in fresh medium the cultures created from the stationary phase cultures had risen on average 61.2% ± 11.0 (± standard deviation, to 1 d.p.) to an average budding index of 74.7% ± 6.2 (± standard deviation, to 1 d.p.) (diagrams 3-3 and 3-4). When starved for a carbon source it has been shown in this investigation that FY1679 falls from an initial budding index of approximately 70% to be stable at about 20% budded cells ("Glucose induced growth arrest", chapter 2). Thus it can be concluded that the cultures used to compare the heat shock resistance of growing and arrested cultures are in either stationary or exponential phase.

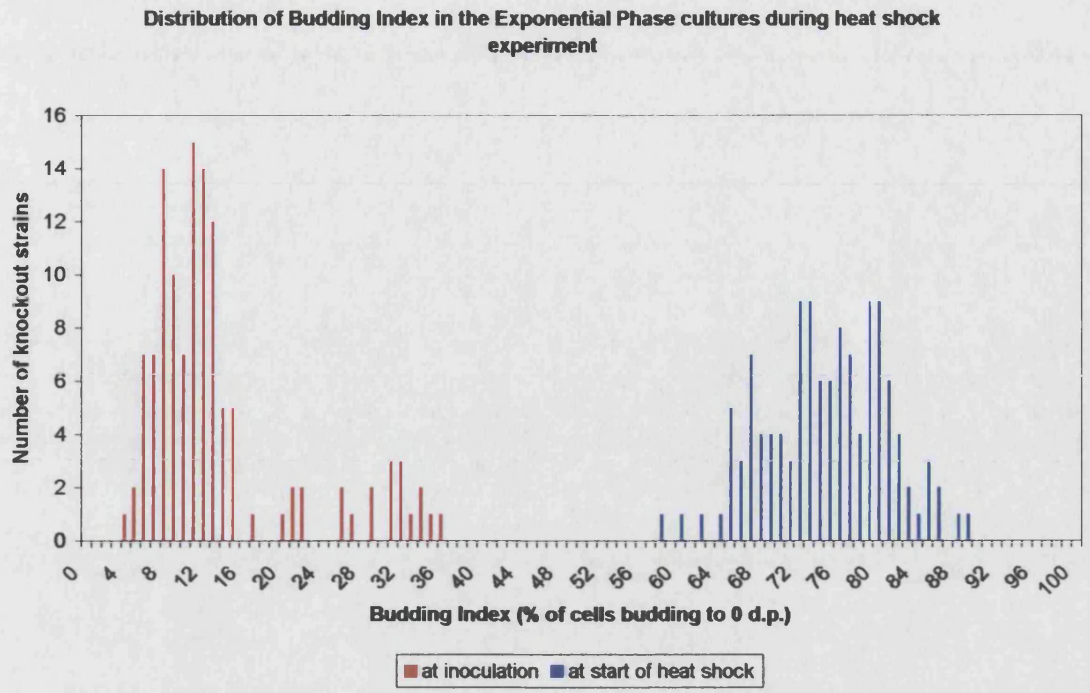


Diagram 3-3: The observed budding index of the exponential and stationary phase cultures.

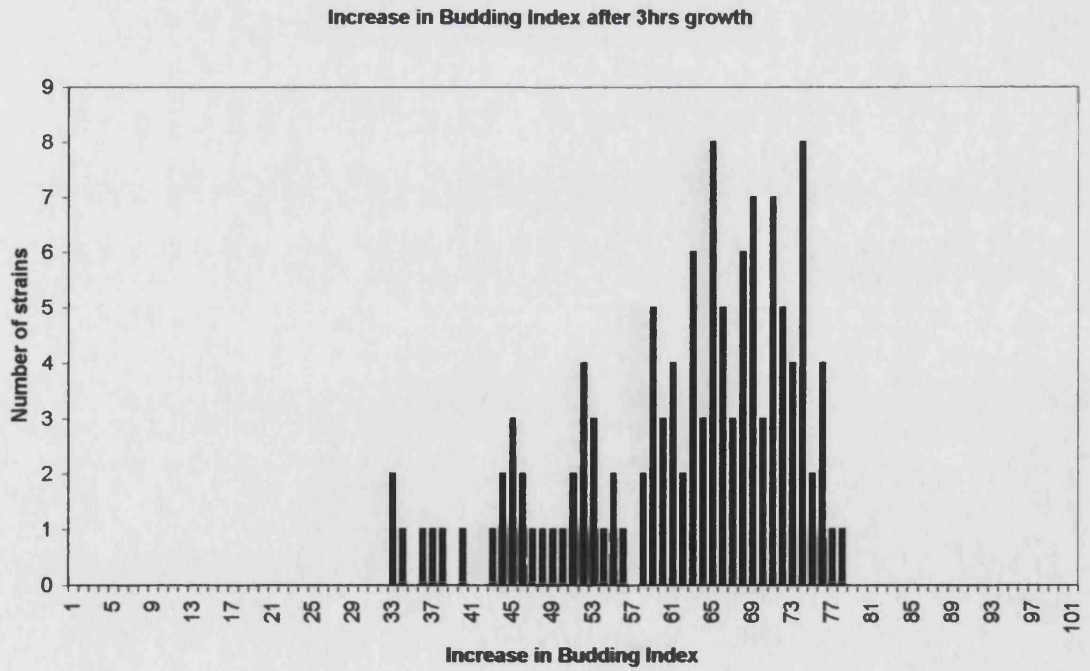


Diagram 3-4: The difference in budding index between stationary phase and exponential phase cultures.

For most of the knockout strains examined there was no difference in the sensitivity to heat shock between the stationary and exponential phase cultures. However in sixteen cases there was an observable difference in the sensitivity of the knockout strains to heat shock. For four strains the exponential phase cultures were more resistant to heat shock than the stationary phase cultures while twelve other knockouts were less resistant to heat shock when in exponential phase compared to stationary phase.

Cells in stationary phase are more heat shock resistant than cells in exponential phase (Werner-Washburne, Braun et al. 1993; Werner-Washburne, Braun et al. 1996) and twelve of the ORF knockouts studied retained this effect. The proteins coded for by these ORFs are:

> FZO1 (mitochondrial biogenesis), COQ4 (ubiquinone biosynthesis), PRO1 (proline biosynthesis), OXA1 (cytochrome oxidase biogenesis), PET309 (stability/translation of COX1 mRNA), YDR523c (serine/threonine protein kinase), PPA2 (mitochondrial inorganic pyrophosphtase), MRPS28 (mitochondrial ribosome small subunit), RPL1B (ribosome large subunit), MRP7 (mitochondrial ribosome protein) , YER087w (has similarity to *E. coli* prolyl-tRNA synthetase) and YLR358c (hypothetical ORF).

There does not seem to be any common theme amongst these ORFs. Most of them are related to mitochondrial activity but since the majority of the ORF knockouts had this feature this result could be reflecting this bias. Nevertheless these genes/proteins are able to develop increased heat shock resistance associated with this transition to stationary phase.

Most of the ORF knockouts revealed no difference between the heat shock sensitivity of the exponential and stationary phase cultures. This indicates that for the majority of the ORF knockout strains examined the heat shock resistance associated with entering stationary phase is not achieved. As we know from examining budding indexes

the cultures are arresting growth so it can be concluded that the majority of the knockout strains studied here are unable to enter stationary phase properly.

The four ORF knockout strains that showed more resistance to heat shock in exponential phase than during stationary phase were *MRF1* (mitochondrial polypeptide chain release factor), *MDH1* (malate dehydrogenase), *ISA1* (required for iron-suphur assembly) and *PHO85* (cyclin dependent protein kinase - may regulate nutrient conditions to the cell cycle). These knockouts that cause hypersensitivity to heat shock do not seem to be functionally related.

Cellular glycogen and trehalose levels of stationary phase cells

Due to time limitations not every gene knockout strain was tested. Therefore a selection was made based on the known characteristics of the gene knockout. Those gene knockouts which are respiratory deficient were not tested because respiratory deficient strains do not accumulate glycogen on entry into stationary phase (Yang, Chun et al. 1998). Forty-five strains had a stationary phase defect when grown on YPD medium but did not have a defect when grown on YPG medium while a further ten gene knockouts were not rescued by growth on YPG medium.

It was decided to study the ten genes whose stationary phase defect could not be rescued by growth on YPG medium. This defect is less dependent on the conditions in which the cells are grown suggesting the genes affected are more essential (Table 3-5).

According to Parrou & François (Parrou and François 1997) the incubation of the cell suspensions in Na_2CO_3 at 95°C destroys all glucose in the medium. Consistent with this, all the untreated preparations contained less than 0.08 $g.l^{-1}$ glucose (this is the minimum detection limit of the glucose testing kit used). Glucose levels in the preparations from *S. cerevisiae* FY1679 treated with amyloglucosidase and trehalase were 0.466 $g.l^{-1}$ and 0.142 $g.l^{-1}$ respectively (Table 3-6). This indicated that glycogen and trehalose levels in stationary phase cells are large enough to be detectable by this method.

Gene KO	Gene function
YDL044c	Mitochondrial protein involved in mRNA splicing and protein synthesis
YDR194c	Mitochondrial RNA helicase of the DEAD box family
YJL188c	Questionable ORF
YJR144w	Mitochondrial genome maintenance protein
YKR006c	Mitochondrial ribosomal protein (*YmL13*)
YKR024c	RNA helicase required for 60S ribosomal subunit assembly
YML061c	DNA helicase involved in mitochondrial DNA repair and telomere length
YMR064w	Nuclear control of ATPase messenger RNA expression protein
YNL139c	Regulatory protein
YNR036c	Strong similarity to ribosomal protein S12

Table 3-5: Gene knockout strains used in this experiment.

Gene knockout	Strain	Amount of glucose in sample ($g.l^{-1}$)		
		Untreated	Amyloglucosidase treated	Trehalase treated
FY1679	diploid	< 0.080	0.466	0.142
YKR024c	MATa	< 0.080	0.086	< 0.080
YKR024c	MATα	< 0.080	< 0.080	0.200

Table 3-6: Level of glucose in tested samples (to 3 d.p.).

In the majority of strains tested the level of glucose in the sample following digestion of glycogen and trehalase was undetectable indicating either very little (less than 0.08 $g.l^{-1}$) or no glycogen and trehalase in the stationary phase cells. The two exceptions to this are shown in Table 3-6 and were contaminated with bacteria that may have caused the unusual results.

The gene knockouts tested do not accumulate glycogen and trehalose upon entry into stationary phase. Whether the lack of these two compounds is the primary cause of the loss of viability in stationary phase or is a secondary effect of the gene knockout is not known. However the lack of two carbohydrate storage compounds, one of which is also a stress protectant, at a time when a cell is preparing for an unknown period of carbon starvation is unlikely to promote long term stationary phase survival.

Appearance of stationary phase cells

When examined under phase-contrast microscopy cells that are in stationary phase appear as refractile or phase bright and are unbudded (Werner-Washburne, Braun et al. 1993) (Diagram 3-5). None of the stains examined showed any visible sign of not being able to enter stationary phase. It can therefore be concluded that these stains are able to detect and respond to stationary phase to a certain degree. If the strains were unable to detect/respond in any way to stationary phase more of the cells would appear to be budded as they attempted to continue through the cell cycle. This suggests that whatever is causing the loss in viability during stationary phase for these strains does not affect the ability of a cell to detect starvation conditions or to respond to the starvation conditions by entering stationary phase.

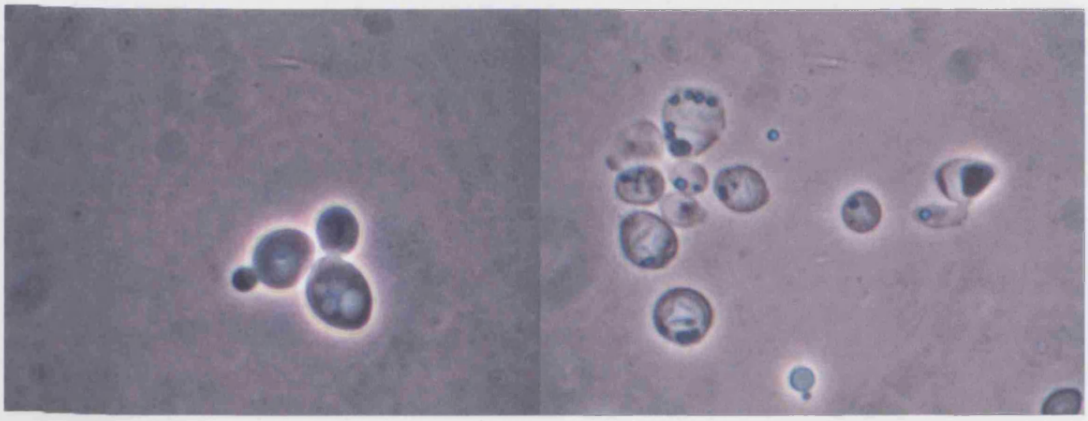


Diagram 3-5: *S. cerevisiae* FY1679 cells in exponential phase (left) and stationary phase (right) at x100 magnification.

Loss of cell viability

Each week after inoculation, for 13 weeks, samples of *S. cerevisiae* ORF knockout cultures were taken, diluted and plated out on solid YPD medium. For each gene knockout strain (each mating type/diploid in both the washed and unwashed cultures) the number of viable cells in the culture was calculated using the formula:

$$\text{Number of viable cells per ml of culture} = \frac{\text{Number of colonies grown on plate}}{\text{Volume (µl) of diluted clture used}} \times \frac{1000}{\text{Dilution Factor}}$$

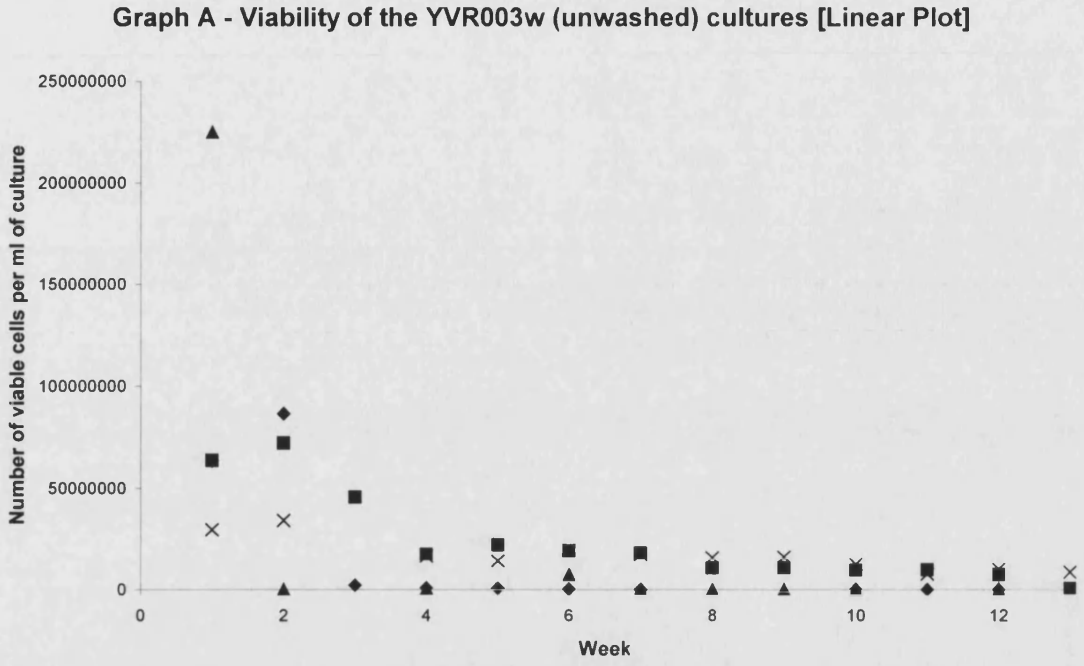
An example of the calculated numbers of viable cells is given in Table 3-7; complete results are in appendix 4.

Week	Unwashed				Washed			
	MATa	MATα	Hetero.	Homo.	MATa	MATα	Hetero.	Homo.
1	$6.35x10^7$	$6.35x10^7$	$2.25x10^8$	$2.96x10^7$	$5.85x10^7$	$3.76x10^7$	$1.68x10^8$	$3.82x10^7$
2	$8.65x10^7$	$7.20x10^7$	0	$3.40x10^7$	$1.00x10^6$	$3.45x10^7$	$1.10x10^7$	$2.85x10^7$
3	$2.00x10^6$	$4.55x10^7$	No data	no data	$2.92x10^7$	$1.95x10^7$	$1.65x10^7$	$2.60x10^7$
4	$8.50x10^5$	$1.75x10^7$	$5.00x10^5$	$1.65x10^7$	$2.00x10^7$	$8.50x10^6$	$1.20x10^7$	$4.00x10^6$
5	$5.00x10^5$	$2.20x10^7$	no data	$1.40x10^7$	$5.00x10^5$	$1.65x10^7$	$9.50x10^6$	$4.00x10^6$
6	$1.00x10^5$	$1.91x10^7$	$7.25x10^6$	$1.96x10^7$	0	$6.30x10^6$	$4.60x10^6$	$4.55x10^6$
7	$4.03x10^4$	$1.80x10^7$	0	$1.71x10^7$	0	$3.55x10^6$	no data	$2.90x10^6$
8	no data	$1.04x10^7$	0	$1.57x10^7$	5	$1.55x10^6$	no data	$3.10x10^6$
9	No data	$1.06x10^7$	500	$1.61x10^7$	0	$1.57x10^6$	$1.10x10^6$	$1.60x10^6$
10	300	$9.35x10^6$	5	$1.22x10^7$	2170	$6.00x10^5$	$1.40x10^6$	$1.65x10^6$
11	45	$9.50x10^6$	no data	$7.50x10^6$	0	$5.65x10^5$	no data	$6.00x10^5$
12	20	$7.15x10^6$	0	$9.85x10^6$	0	no data	no data	$5.55x10^6$
13	no data	$5.00x10^5$	no data	$8.40x10^6$	0	$1.40x10^5$	no data	$2.35x10^5$

Table 3-7: Number of viable cells per ml of culture of the gene knockout YCR003w (numbers given to 2 decimal places).

An exponential growth/decay curve assumes a constant probability of growth/death over a defined period of time. A cell culture following an exponential growth/decay follows the formula: $N_t = N_0.e^{Kt}$. Where N_t is the number of cells after time = t, N_0 is the number of cells at time = 0, K is the decay constant and t is the time. K will be positive for an exponential growth, and negative for an exponential decay. In this form the graph of the growth/decay will follow an exponential curve. The formula can be rearranged into the form: $Ln(N_t) = Ln(N_0) + Kt$. In this form the data will produce a straight-line graph of the type y = mx + c. where the natural log of the number of cells is plotted on the y-axis (y) and the time elapsed is plotted on the x-axis (x). The natural log of the number of cells at time = 0 is c, and the decay constant, K, is the gradient of the line (m). When these results of this experiment are plotted on a graph (viable cells per ml of culture versus week) the results appear to follow an exponential decay. This appears to be confirmed when a semi-logarithmic plot of the results is drawn (Ln(viable cells per ml of culture) versus week) (example, see Diagram 3-6).

All the data series in this experiment appear to form a straight line when plotted on a semi-log graph. This was tested by calculating the coefficient of determination (r^2-values) for each of the data sets. The r^2-values vary between 1 and 0; the value 1 indicating that the data fitted a straight line perfectly and the value 0 that the data points were completely random. If the r^2-values are calculated for the strains/conditions used in this experiment it is seen that most of the data follow closely a straight line (Table 3-8) indicating that the strains being studied in this experiment are following an exponential decay in the loss of cell viability.



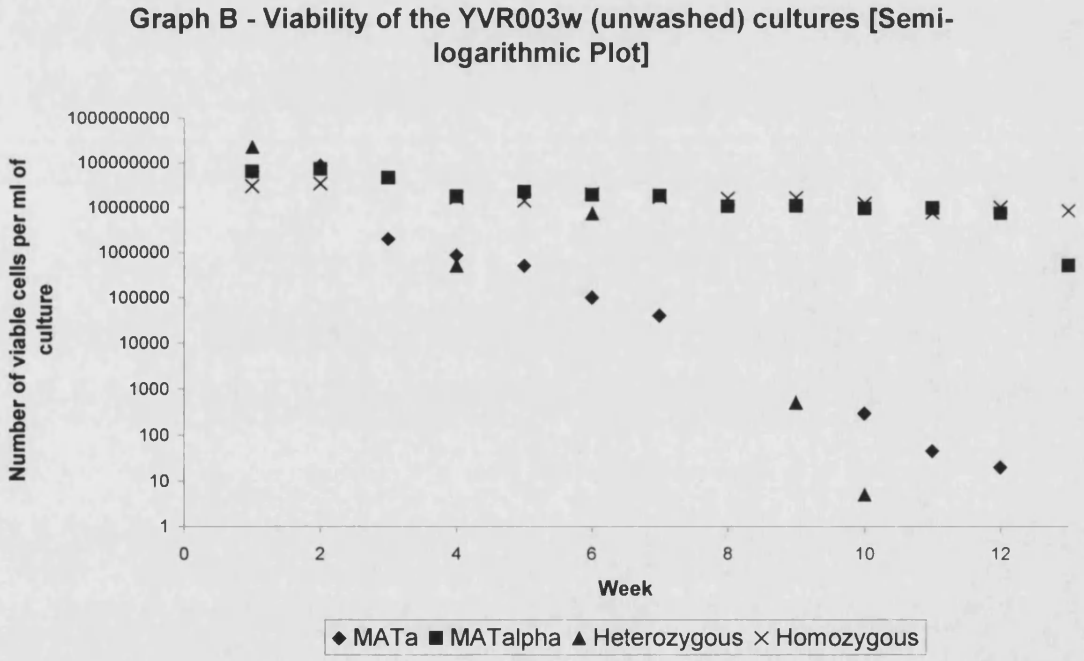


Diagram 3-6: Viability of the YCR003w (unwashed) culture plotted on linear (A) and semi-logarithmic (B) plots. On both graphs the MATa strain is plotted as a diamond, the MATα strain as a square, the heterozygous strain as a triangle and the homozygous strain as a cross.

ORF knockout strain	Unwashed				Washed			
	MATa	MATα	Homo. diploid	Hetero. diploid	MATa	MATα	Homo. diploid	Hetero. diploid
YCR003w	0.99	0.76	0.81	0.85	0.68	0.97	0.73	0.83
YDL202w	0.99	1.00	0.98	0.59	0.99	1.00	0.94	0.43
YDL405w	0.94	0.90	0.90	0.87	0.36	0.93	0.96	0.91
YGL135c	0.85	1.000	0.53	0.78	0.85	0.95	0.80	0.72
YGL220c	0.90	0.89	0.93	0.91	0.70	0.65	0.94	0.86
YDR194c				0.02				0.06
YJR113c	0.96	0.97	0.95	0.89	1.00	0.98	0.59	0.94
YKR085c	0.98	0.93	0.92	0.90	0.92	0.82	0.37	0.73
YNL177c	0.83	0.63	0.90	0.66	0.70	0.56	0.80	0.65
FY1679	0.72				0.81			

Table 3-8: Calculated r^2 values (to 2 d.p.).

The decay constant for each knockout strain is shown in Table 3-9. The cell types that show a stationary phase defect (MAT**a**, MATα, and homozygous diploid) all have average K-values of approximately –0.7 to –1.0 while the heterozygous diploid and the strain FY1679 (which do not lose viability in stationary phase) have K-values of approximately –0.3 to –0.4. The steeper gradients of the haploid and homozygous diploid types show that these cultures lose cell viability at a faster rate than the heterozygous diploid and FY1679. This suggests that they could both be losing viability by the same mechanism but the process is accelerated in the MAT**a**, MATα and homozygous diploid types.

Culture type	KO strain	MATa	MATα	Heterozygous	Homozygous	FY1679
Unwashed	YCR003w	-1.4	-0.3	-1.8	-0.1	-0.3
	YDL202w	-1.3	-1.2	-0.3	-1.2	
	YDL405w	-0.2	-0.3	-0.2	-1.2	
	YGL135c	-0.3	-0.6	-0.1	-0.2	
	YGR220	-0.3	-0.3	-0.2	-0.3	
	YDR194	No data	No data	-0.0	No data	
	YJR113	-1.1	-1.4	-0.3	-1.0	
	YKR085	-1.0	-1.1	-0.3	-0.9	
	YNL177c	-1.1	-0.3	-0.2	-0.5	
Washed	YCR003w	-1.5	-0.5	-0.4	-0.3	-0.4
	YDL202w	-1.3	-1.3	-0.3	-1.3	
	YDL405w	-0.2	-0.4	-0.3	-1.4	
	YGL135c	-0.5	-0.4	-0.3	-0.3	
	YGR220	-0.3	-0.3	-0.3	-0.4	
	YDR194	No data	No data	-0.07003	No data	
	YJR113	-2.1	-1.7	-0.5	-1.2	
	YKR085	-2.2	-1.6	-0.4	-0.4	
	YNL177c	-0.5	-0.8	-0.4	-0.4	
Average (unwashed)		-0.8	-0.7	-0.4	-0.7	-0.3
Average (washed)		-1.1	-0.9	-0.3	-0.7	-0.4
Average (all)		-1.0	-0.8	-0.4	-0.7	-0.4

Table 3-9: Decay constant (gradient of the line of best fit) of cell culture viability (to 1d.p.).

If the loss in viability of each of the cultures is affected by either remaining in spent medium or by being stored in sterile water the decay constant for the "unwashed" cultures will be different from the "washed " cultures. This can be tested by using a paired t-test on the data. The null hypothesis (H_0) for the paired t-test is:

The average value for the decay constant for the unwashed cultures is equal to the average value for the decay constant of the washed cultures.

The value of the test statistic is calculated using the formula:

$$t = \frac{\bar{d}}{s_d \div \sqrt{n}} \quad \text{(with n-1 degrees of freedom)}$$

Where $\bar{d}$ is the average of $K_{unwashed} - K_{washed}$, s_d is the standard deviation of the unwashed minus the washed values, and n is the number of values. The calculated value of t, at the appropriate degrees of freedom can be used to obtain a P-value, P(α). If P(α) is less than 0.05 then the null hypothesis is rejected at the 5% significance level. The values of P(α) for each of the 4 different ploidy/mating types is shown in Table 3-10.

Ploidy/mating type	MAT**a**	MATα	Heterozygous	Homozygous
P(α)	0.3	0.1	0.8	0.7

Table 3-10: Values of P(α) for the paired t-test comparing unwashed to washed conditions (to 1 d.p.).

For each of the four ploidy/mating types the calculated value of P(α) is greater than 0.05. Therefore the null hypothesis is not rejected, and it can be concluded that there is no difference between the decay constants for the unwashed and washed cultures. This means that it must be concluded that the loss in cell viability for the ORF knockouts tested in

spent medium proceeds at the same rate as for cells in sterile water. This is in contrast to previous studies that have shown that if cells in stationary phase are in water instead of spent medium they have prolonged stationary phase survival (Granot and Snyder 1991; Granot and Snyder 1993; MacLean, Harris et al. 2001). It can also be seen however that the unmodified *S. cerevisiae* strain, FY1679, used as a control, also loses cell viability at a faster rate when incubated during stationary phase in water.

Previous studies did not use the strain FY1679, which could explain the difference. However FY1679 and strains used in the previous studies are both similar to the strain S288C. Also previous studies only considered the changes in culture viability to a point that is equivalent to approximately four weeks post inoculation. If the data generated for this study for only the first four weeks, instead of all 13, is considered the conclusions change slightly.

The control strain (*S. cerevisiae* FY1679) retained cell viability to a greater degree when incubated in water for a period of 4 weeks, with this situation reversing after 13 weeks (Table 3-11). This shows that while incubation in water does reduce the loss of viability over a shorter period this ability of water to reduce the loss in viability is not effective for longer periods of time. The change in the loss of viability of the cell cultures of the ORF knockouts is not as obvious but there is a trend towards stationary phase cultures incubated in water having a lower rate of loss of viability than those cultures incubated in spent medium. This can be seen in Table 3-12. This indicates a greater occurrence of incubation of stationary phase cells in water resulting in a slower loss of viability.

Culture	Weeks 1 - 4	Weeks 1 - 13
Unwashed	-0.4	-0.3
Washed	0.0	-0.4

Table 3-11: Decay constant (gradient of the line of best fit) of cell culture viability for cultures of *S. cerevisiae* FY1679 (to 1 d.p.).

Weeks 1 - 13				Culture	Weeks 1 - 4			
MATa	MATα	Homo.	Hetero.		MATa	MATα	Homo.	Hetero.
0.1	0.2	0.2	-1.4	YCR003w	-1.7	0.1	0.5	-1.3
0.1	0.1	0.1	-0.0	YDL202w	0.9	0.6	1.0	0.1
-0.0	0.1	0.2	0.2	YDL405w	0.3	-0.1	-0.9	0.1
0.2	-0.2	0.1	0.1	YGL135c	0.1	0.8	0.2	0.3
0.0	0.0	0.1	0.1	YGL220c	-0.2	-0.2	-0.0	-0.0
No data	No data	No data	0.0	YDR194c	No data	No data	No data	-0.1
1.0	0.3	0.2	0.2	YJR113c	1.6	1.7	0.9	0.0
1.2	0.5	-0.8	0.	YKR085c	No data	1.5	1.0	-0.1
-0.6	0.5	-0.1	0.	YNL177c	1.	-0.6	-0.0	-0.1

Table 3-12: Decay constants of unwashed cultures minus decay constants of washed cultures. Shading indicates a positive figure (white), negative figure (black) or no data (grey) (to 1d.p.).

Discussion

Visual appearance of stationary phase cells

None of the ORF knockouts in stationary phase that were examined showed any obvious difference from the appearance of *S. cerevisiae* FY1679 in stationary phase. This suggests that all these ORF knockout strains are able to recognise and react to starvation conditions

even if they cannot maintain viability for a long period of time under these conditions. Therefore, especially for those ORF knockouts interfering with signalling pathways, *S. cerevisiae* must have more than one pathway capable of inducing stationary phase that prevents any one single ORF knockout/mutation from preventing entry into stationary phase. This may have evolved because of the evolutionary pressure exerted by starvation, a very common stress for *S. cerevisiae* in the wild (Werner-Washburne, Braun et al. 1993). The ability of the yeast to survive would be made very much more difficult if single mutations could destroy the ability of a cell to react to starvation conditions.

Kinetics of loss of viability in stationary phase

Over the long term, the knockout strains that were tested lost viability at approximately the same rate in spent medium as in sterile water. However during short-term incubation cultures retained viability for longer when incubated in water during stationary phase, which has been observed in other experiments (Granot and Snyder 1991; Granot and Snyder 1993; MacLean, Harris et al. 2001). These previous studies showed that incubation with a utilisable carbon source results in loss of viability of a stationary phase culture whereas incubation in water, a non-utilisable carbon source or a nitrogen source does not result in a loss of viability. This suggests that the presence of the utilisable carbon source stimulates cell growth and that the loss of viability is caused by the cells starving due to the lack of any other nutrient in the medium.

A possible explanation for the difference between short- and long-term incubation in water could be that during short-term incubation non-utilisable carbon and nitrogen sources in spent medium prevents stationary phase from being fully established, unlike cells incubated in water resulting in a faster loss of viability for cultures maintained in spent medium. Over the longer-term however, there is a change that allows the cultures in spent medium to enter fully stationary phase which causes the loss of viability of cells in spent medium to drop to the same levels as the cells that are incubated in water. The

change that causes this could be one of a number of things, including changes of metabolite levels in the medium or the cell itself.

Cellular glycogen and trehalose levels

The ORF knockouts studied cannot accumulate glycogen and trehalose in stationary phase. This could be the reason that these strains lose viability in stationary phase. Trehalose provides heat shock protection and glucose provides starvation protection when as low as 1.5% (w/w) of dry cell mass, with larger amounts of trehalose leading to longer lasting protection (Plourde-Owobi, Durner et al. 2000). The importance of the levels of trehalose and glycogen could be tested by artificially increasing the level of these two carbohydrates in the cell without using the cellular metabolism. The intracellular levels of trehalose can be increased by incubation in a medium that contains trehalose (Plourde-Owobi, Durner et al. 2000). However the presence of the carbohydrates in the medium would probably cause cells to exit stationary phase and resume growth. A possible way of avoiding these would be to use a conditional mutant of a component of either the RAS/cAMP or TOR pathways since a mutation that inactivates either of these pathways results in constitutive stationary phase arrest (Herman 2002).

The rest of the ORF knockouts that have been shown to have a stationary phase defect should have their stationary phase levels of glycogen and trehalose determined. This would reveal whether all the stationary phase essential ORFs, when knocked out, have reduced levels of these two carbohydrates. If they do not all have reduced glycogen and trehalose levels it proves that these two chemicals are not essential for stationary phase survival. If all the ORF knockouts were found to have reduced levels then it would reinforce the importance of intracellular glycogen and trehalose levels for long-term stationary phase viability. If it is proved that all stationary phase essential ORF knockouts have reduced glycogen and trehalose levels in stationary phase, ORFs that are not essential for stationary phase should be studied. If any ORFs not essential for stationary phase could

be proved to have reduced glycogen and trehalose levels it would show that intracellular levels of the carbohydrates is not essential for stationary phase and that the observed low levels are only a secondary effect.

Heat shock resistance

Most strains appear to be affected in their ability to resist heat shock, only twelve of the strains tested retained a greater resistance to heat shock when in stationary phase compared to when they are in exponential phase. Whether this is the reason that the cells lose viability in stationary phase, or is a secondary effect of the ORF knockout, cannot be concluded from the data collected in this investigation. The study of the trehalose levels of the ORF knockouts with a stationary phase defect suggests that the increase in heat shock sensitivity is a secondary affect of the ORF knockout. Trehalose is known to protect cells against heat shock and starvation (Plourde-Owobi, Durner et al. 2000), and all of the ORFs that were found to have reduced intracellular levels of trehalose were found to be more sensitive to heat shock in stationary phase than the wild type.

As has already been mentioned the remaining ORF knockout strains that have been found to be essential for stationary phase should have glycogen and trehalose levels measured. As it is possible that the stationary phase heat shock sensitivity of the strains studied in this investigation is due to the lack of trehalose in stationary phase cells it would be interesting to find out what the stationary phase levels of trehalose in *MRF1*, *MDH1*, *ISA1* and *PHO85* knockout strains are. These strains were observed to have, like the strain with no ORF knockouts, a greater resistance to heat shock in stationary phase. If they have normal stationary phase levels of trehalose it could be concluded that the loss of stationary phase heat shock resistance in the ORF knockout studied is most probably directly related to the levels of trehalose in the cells.

Chapter 4

Bioinformatics

Summary

A bioinformatics approach was used to gain insights into the nature of the genes which, when knocked out, cause a stationary phase defect. An analysis of the transcription profiles of all these genes suggested roles for several proteins currently of unknown function. The Pathway Analysis program that also uses transcription profiles generates a group of co-regulated genes that could be required for stationary phase. Like the set of genes identified as being essential for stationary phase, this group of co-regulated genes has a mitochondrial bias. Pattern discovery and multiple sequence alignment programs were used to generate four potential stationary phase associated upstream regulatory sequences. Known upstream regulatory sequences for transcription factors associated with stationary phase were shown to not be over represented in the upstream regions of stationary phase essential genes.

The use of bioinformatics in research

In recent years the development of high-throughput screening techniques has led to an explosion in the amount of data available to yeast researchers. In 1996 there were approximately $1x10^6$ DNA sequences in the GenBank database. Three years later this number had risen to $5x10^6$ (Tamames and Tramontano 2000), and by January 2003 to $22.3x10^6$ sequences (http://www.ncbi.nlm.nih.gov/Genbank/GenbankOverview.html). This large volume of data can make it difficult to locate specific information and identify patterns. For many researchers, including those studying *S. cerevisiae*, there is too much information on the function, structure and expression of genes, genomes, proteomes and transcriptomes to manually analyse all the data. Consequently bioinformatics techniques

have evolved to aid the researcher in sorting and using the information available to them. The Internet is instrumental in this by providing easy and quick access to both large amounts of data and to programs with which to interpret the data on a global scale.

These bioinformatic/data analysis applications can be a set of very powerful tools allowing similarities to be extracted and predictions made. However, it must be remembered that the output from these analyses are usually predictions and if the data are of poor quality or erroneous or the wrong settings are used in the software then the predictions generated will be inaccurate or misleading. Once a prediction has been made "wet-lab" experiments should be conducted to test it.

Bioinformatics research conducted in this investigation concentrated mainly on DNA and transcription. This is because there are more tools and data available to study these areas than there are tools and data for investigating proteomics. Other important factors to consider are how efficiently the mRNA that is produced is translated and how stable the resultant proteins are in the stationary phase cell. Studying transcription in stationary phase cells is less informative than the same studies in exponentially growing cells. This is because there is very little active transcription and the efficiency of translation of the mRNAs changes in stationary phase cells compared to growing cells (Werner-Washburne, Braun et al. 1993; de la Cruz, Prieto et al. 2002; Cyrne, Martins et al. 2003). Studying transcription is not completely without merit. As there is almost no transcription and translation in stationary phase cells most of those proteins that are required for stationary phase must be produced before the cell enters stationary phase. The genes coding for these proteins will therefore be likely to be upregulated in growing cells that are preparing to enter stationary phase.

Materials and Methods

Downloading/Generation of data - Upstream Sequences

The upstream sequences of the 105 genes identified as causing a stationary phase defect were extracted from the YEAST_-1000_+2_W_all.fa set of sequences downloaded using the GENOME tool at http://ep.ebi.ac.uk/. For each gene a 1003 bp sequence was downloaded, from the ATG start codon to 1000 bp upstream of the A in the start codon. This set of sequences is called YPD105. Thirteen of these strains however show a stationary phase defective phenotype when grown in SC medium, but do not show this defect when grown in YPD medium. Removing these gene knockouts from the list left 92 gene knockout strains that have a stationary phase defect when grown in YPD medium. The YPD105 set of gene was edited to remove those genes that do not give a stationary phase defect when the knockout strain is grown in YPD medium. This set is called YPD92. A third set of gene sequences of 100 randomly selected genes (RANDOM100) was also created using the GENOME tool (http://ep.ebi.ac.uk/EP/GENOMES/).

Downloading/Generation of data - Transcription Profiles

The results of three transcription profile experiments were downloaded from the Stanford Genomic Resources website (http://genome-www4.stanford.edu/cgi-bin/SGR/publication/publicRef?organism=S.+cerevisiae) and were called Eisen (Eisen, Spellman et al. 1998), Ferea (Ferea, Botstein et al. 1999) and Gasch (Gasch, Spellman et al. 2000). Using Microsoft Excel the transcription profiles of the 92 genes with a stationary phase defect when grown in YPD medium were extracted from each of the downloaded datasets. The full set of 105 genes was not used, as 13 of the genes do not show a stationary phase defect in YPD unlike the other 92 genes. This difference in the phenotypes of the genes could result in a different pattern of transcriptional regulation. Therefore the slightly smaller set of genes that all have a stationary phase defect when grown in YPD medium

was used, and those genes that have a defect when grown SC, but not YPD, medium were excluded.

Conditions not relevant to stationary phase/nutrient starvation were deleted from each of the datasets and then the datasets were combined (table 4-1). The conditions and time courses from the retained datasets that were not deleted were chosen because they were either monitoring gene expression during the growth from fresh medium (through the diauxic shift) into stationary phase. Which is where gene expression for the adaptation of the cell from growth to stationary phase will be occurring. Or are time courses monitoring other nutrient depletion processes (such as nitrogen starvation and sporulation). As these processes result in growth arrest from the absence of a nutrient and use many of the same signalling pathways transcription regulation is very likely to be very similar as well.

With the conditions listed in table 4-1, the genes were ordered according to their transcription profiles using the program Cluster (Eisen, Spellman et al. 1998). The ORFs were first ordered using the Self Organising Map feature of Cluster (Using the settings: Xdim = 1, Ydim = 12 and Iterations = 100000). The output order of the ORFs from this was used to create the values for GORDER for the original dataset. A hierarchical clustering of the ORFs was then completed using Spearman Rank Correlation and complete linkage clustering. A non-parametric Pearson correlation coefficient was used as there were some missing values in the datasets. The results of the ordering by the program Cluster were then visualised using the program Treeview (Eisen, Spellman et al. 1998).

Dataset	Conditions used from the dataset	
	Time course	**Time points**
Eisen	Sporulation	spo 0, spo 2, spo 5, spo 7, spo 9, spo 11, spo5 2, spo5 7, spo5 11, spo-early, spo mid
	Diauxic Shift	diau a, diau b, diau c, diau d, diau e, diau f, diau g
Ferea		Evolved strain 1 vs evolved strain 1, Parental vs. evolved strain 1, Parental vs. evolved strain 2, Parental vs. evolved strain 3
Gasch	Amino Acid starvation	0.5 hour, 1 hour, 2 hour, 4 hour, 6 hour
	Nitrogen depletion	0.5 hour, 1 hour, 2 hour, 4 hour, 8 hour, 12 hour, 1 day, 2 days, 3 days, 5 days
	Diauxic shift	0 hour, 9.5 hour, 11.5, 13.5 hour, 15.5 hour, 18.5 hour, 20.5 hour
	Growth in YPD	2 hour, 4 hour, 6 hour, 8 hour, 10 hour, 12 hour, 1 day, 2 days, 3 days, 5 days
	YPD stationary phase	2 hour, 4 hour, 8 hour, 12 hour, 1 day, 2 days, 3 days, 5 days, 7 days, 13 days, 22 days, 28 days

Table 4-1: profile conditions used from each dataset.

Known Upstream Regulatory Sequences

There are many regulatory proteins and Upstream Activation Sequences (UAS) that may have some effect on the regulation of stationary phase genes. These include:

Gln3p; This protein is used in carbon and nitrogen source regulation (Bertram, Choi et al. 2000; Bertram, Choi et al. 2002). There is currently no consensus binding sequence known for this protein. However Gln3p is known to bind to sites upstream of the

ORFs YDL215c, YGR019w and YPR035w (SCPD 1998). A possible consensus sequence can be determined by comparing these sites (table 4-2). Adr1p; Involved in carbohydrate (glycerol) metabolism, the consensus binding site is 5'-TCTCC-3' (SCPD 1998). Mig1; Mig1p is involved in glucose repression (it represses in the presence of glucose). The consensus binding site for Mig1p is 5'-CCCCRNNWWWWW-3' (SCPD 1998). Swi5p; Possibly involved in the cell cycle With consensus sequence 5'-KGCTGR-3' (SCPD 1998).

YDL215c	5'	CTAATCTAATC	3'
YGR019w	5'	TAATCTAATC	3'
YPR035w	5'	GATAAGATAAG	3'
Common	**5'**	nn**TAA**nn**TAA**nn	**3'**

Table 4-2: Identifying bases common to known Gln3p binding sites (SCPD 1998).

STRE; One of the proteins to bind to the STRE (stress response) element (5'-CCCCT-3') is Msn2/4 (Ruis and Schüller 1995). STRE regulated genes are known to be upregulated after the diauxic shift (Werner-Washburne, Braun et al. 1996). CRE; cAMP responsive Element 5'-TGACGTCA-3' (Gancedo 1998; Gachon, Gaudray et al. 2001). PDS; Gis1p is known to bind to PDS (Post Diauxic Shift Element). This site in YBL075c has the sequence 5' TTAGGGAT-3' (SCPD 1998). CSRE; Sip4p is activated by the Snf1 complex and binds to CSRE (Carbon Source Response Element). Cat8p possibly binds to CSRE in response to Snf1 activation (Carlson 1999; Roth and Schüller 2001). The consensus sequence for CSRE is 5'-YCGGAYRRAWGG-3' (SCPD 1998).

There are more proteins and UAS's than those listed here that are known to be involved in gene regulation due to starvation conditions. These include UAS_{HSF} (Tamai,

Liu et al. 1994), GCS1 (Ireland, Johnston et al. 1994) and Rpi1 (Sobering, Jung et al. 2002). However the binding sequences for these proteins and UAS have not yet been identified.

Results

Gene Grouping According to Transcription Profiles

Large-scale analyses of the yeast transcriptome are now available. These studies provide a large amount of data about the transcription of almost every gene in the genome under different conditions. Using the program Cluster (Eisen, Spellman et al. 1998) the genes was ordered according to their transcript profiles. Once Cluster has finished grouping genes with similar transcript profiles the program Treeview (Eisen, Spellman et al. 1998) was used to visualise the result. The diagram drawn by Treeview shows a different gene on each line, and a different condition in each column. The red & green blocks indicate an increase or decrease respectively in mRNA levels for that gene under each condition (with black indicating no change and grey no data), with the intensity of the colour indicating the magnitude of change. By examining where genes of unknown function occur it is possible that clues to the function of the gene can be found. The 92 ORFs that are essential for the maintenance of viability in stationary phase, organised by their transcript profiles (using the program cluster), are shown in diagram 4-1.

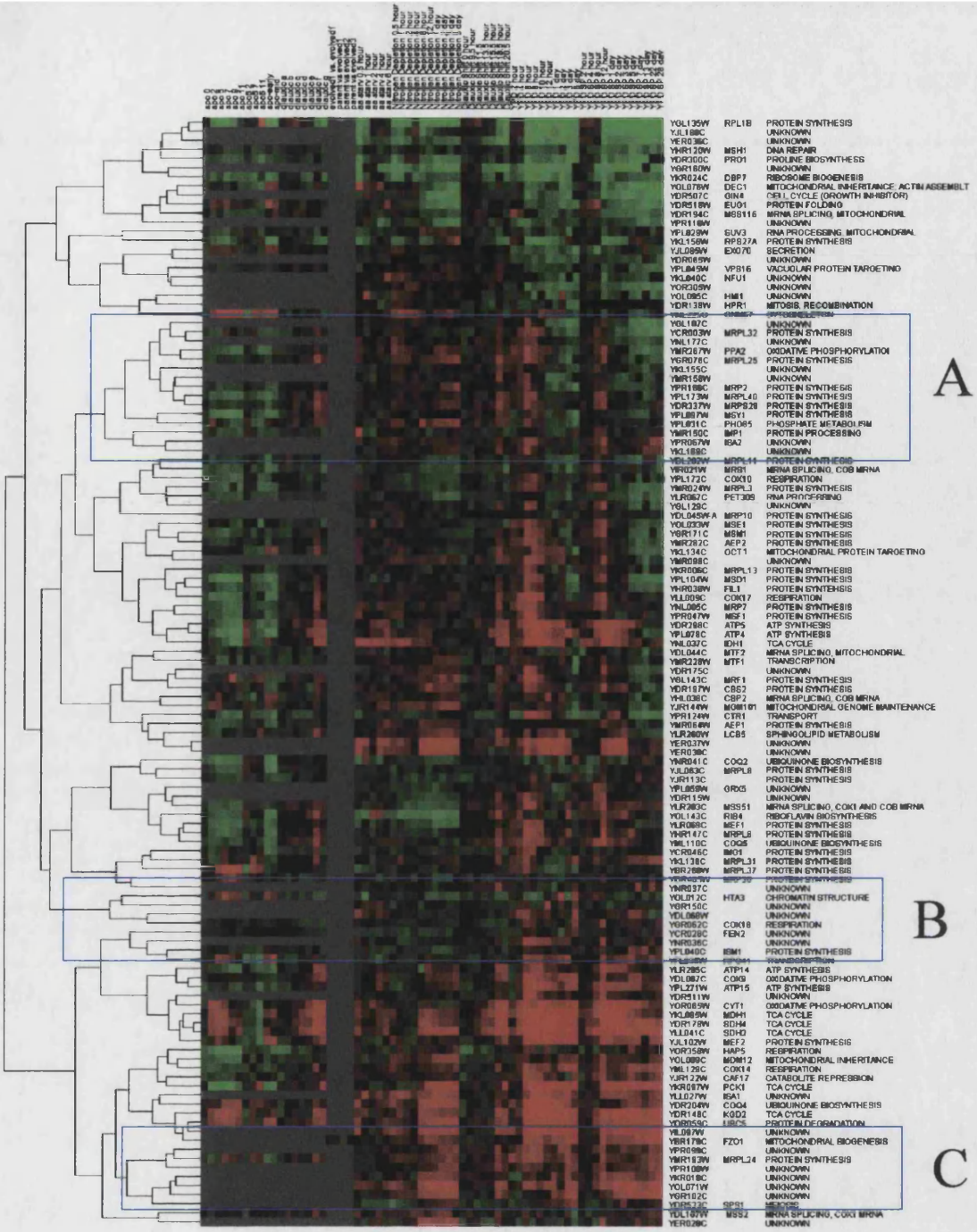


Diagram 4-1: Complete linkage clustering using Spearman rank correlation of the 92 stationary phase essential ORFs.

The unknown ORFs in highlighted area "A" (diagram 4-1) have been ordered alongside ORFs that have been identified as being related to protein synthesis. This therefore suggests that these unknown ORFs have a function that is also related to protein synthesis. In the highlighted areas "B" and "C" (diagram 4-1) several ORFs of unknown function are grouped together. Although this does not reveal what the functions of these

ORFs are, since they are regulated in a similar manner it suggests that the ORFs in each group have a similar function to each other.

Pathway Analysis program

The Pathway Analysis program (Ihmels, Friedlander et al. 2002) also uses transcript profiles to group genes. The program Cluster takes all the inputted genes and groups them into one group based upon the transcript profiles. The Pathway Analysis program however takes the input set of genes and locates those conditions from the transcript profile that best characterises the inputted set of genes. It then recovers from the whole genome those genes that best fit the chosen conditions. This has the advantage of being able to discard genes that have been mistakenly added to the input set (e.g. false positives from a screen) while at the same time being able to retrieve those genes that were not in the original group but whose transcript profiles match the input set of genes to a high degree (false negatives). If the set of genes submitted contains a set of co-regulated genes the transcript profile will remain basically the same when random genes are added it resulting in a high percentage of overlap between the transcript profile of the original set and the set with the randomly added genes. If there are no co-regulated genes in the submitted set any number of randomly added genes will result in a large difference between the two transcript profiles.

When the 92 genes that cause a stationary phase defect when grown in YPD medium were used in the Pathway Analysis program a mitochondrial set of genes was returned. If the recurrence profile of the YPD92 set of genes is examined (Diagram 4-2) it can be seen that the set returns a profile that is significantly different from a control set of random genes. It also retains a similar transcription profile after a large number of random genes (relative to the number of genes in the submitted set) have been added. This suggests that the submitted set of genes consists of a large fraction of co-regulated genes. This result

was expected because an inspection of the input set of genes showed that it has a relatively high number of genes that were related to the mitochondrion and its functions.

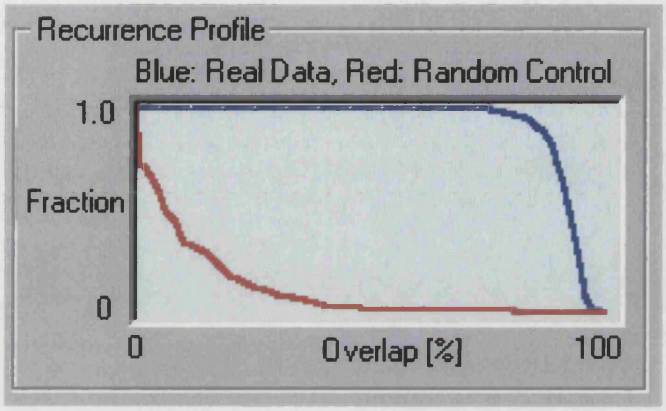


Diagram 4-2: Recurrence profile. This plots the percentage overlap between the transcription profile of the submitted set of genes with the profile of a set made of the submitted genes with a number of random genes versus the fraction of the second set of genes that is made from random genes.

This result suggested that the mitochondrion, or its functions, is very important to the maintenance of viability in stationary phase. The importance of the mitochondrion can be seen not only from the group of mitochondrial genes returned by the program but also by the mitochondrial bias in the input set of genes.

Multiple sequence alignment

The 1 Kb upstream sequences of the YPD92 set of sequences were aligned using the multiple sequence alignment program MAGI, available on the HGMP website (http://www.hgmp.mrc.ac.uk). Default settings were used, with the exception of the Gap start penalty that was set to a value of 2.0. Using the default gap start penalty of 15, the ATG start codon was not located. As every sequence has this codon at the 3'-terminal end

of the sequence it should appear at the terminal end of the consensus sequence as well. Thus the gap start penalty was reduced until the start codon appeared in the consensus sequence. The consensus sequence generated by this alignment located the 3'-terminal ATG present in all the sequences. Other than the start codon only scattered bases appeared on the consensus sequence. Therefore no possible regulatory sequences can be identified. When this was repeated with the sequences for the 105 genes identified with stationary phase defects as before only scattered consensus bases were found, but the ATG start codon was not identified. A control using 100 random genes located the ATG start codon and no obvious upstream sequences.

Search for binding sites for transcription factors associated with stationary phase

To search for known patterns of bases in the *Saccharomyces cerevisiae* genome the Genomic Pattern Matching program on the RSA-tools website was used (http://www.flychip.org.uk/rsa-tools/) (van Helden, André et al. 2000). The sequences that were searched for are shown in table 4-3 and the search options that were used were:

Organism: *Saccharomyces cerevisiae*

Sequence type: *upstream*

From: *-1000*

To: *-1*

Allow overlap with upstream ORFs: *Yes*

Search strands: *both strands*

Prevent overlapping matches: *Yes*

Return: *match positions*

flanking residues: *4*

Origin: *End*

Format: *Table*

Substitutions: *0*

Sites		Consensus sequence
Transcription factor binding sites	Gln3p	TAANNTAA
	Adr1p	TCTCC
	Mig1p	CCCCRNNWWWWW
	Swi5p	KGCTGR
Upstream Activation sequences	STRE	CCCT
	CRE	TGAACGTCA
	PDS	TTAGGGAT
	CSRE	YCGGAYRRAWGG

Table 4-3: Sites searched for

The RSA-tools search program returned a text file with the results in it. These results were then imported into Excel and the numbers of matches for each pattern in the upstream region of the ORFs were counted. A list of ORFs was downloaded from the Stanford website (http://www.yeastgenome.org/gene_list.shtml), and this was used to determine the distribution of the matches across the whole genome (table 4-4). From this distribution the matches to only those ORFs that cause a stationary phase defect when knocked out and the strains were grown in YPD medium (Appendix 3; *Stationary phase essential genes (after secondary screen; SGDP set of ORF knockouts)*) were extracted (table 4-5).

Number of hits	Binding Sites				Upstream Activation Sequences			
	Adr1p	Gln3p	Mig1p	Swi5p	CRE	CSRE	STRE	PDS
0	2769	3912	5820	3110	6148	6165	4187	6104
1	2108	1641	331	2083	19	2	1517	63
2	876	479	15	733	0	0	363	0
3	306	111	1	183	0	0	69	0
4	80	18	0	44	0	0	25	0
5	27	2	0	9	0	0	5	0
6	0	1	0	3	0	0	1	0
7	1	0	0	1	0	0	0	0
8	0	0	0	1	0	0	0	0
9	0	1	0	0	0	0	0	0
10	0	2	0	0	0	0	0	0
Total	6167	6167	6167	6167	6167	6167	6167	6167
Mean	0.8	0.5	0.1	0.7	0.0	0.0	0.4	0.0
$SD_{population}$	1.0	0.8	0.2	0.9	0.1	0.0	0.7	0.1

Table 4-4: The number of ORFs found with "x" number of binding sites within 1Kb upstream of the transcription start. The standard deviation of the population ($SD_{population}$) was used instead of standard deviation, as the distribution of matches through the entire genome (not a sample from it) was being studied. Mean and standard deviation are given to 1 d.p.

Number of hits	Binding Sites				Upstream Activation Sequences			
	Adr1p	Gln3p	Mig1p	Swi5p	CRE	CSRE	PDS	STRE
0	42	49	87	38	91	92	91	68
1	35	37	4	35	1	0	1	21
2	12	5	0	14	0	0	0	3
3	3	0	1	3	0	0	0	0
4	0	0	0	2	0	0	0	0
5	0	1	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0
Total	92	92	92	92	92	92	92	92
Mean	0.7	0.6	0.1	0.9	0.0	0	0.0	0.3
Standard Deviation	0.8	0.8	0.4	0.9	0.1	0	0.1	0.5

Table 4-5: The number of stationary phase essential ORFs found with "x" number of binding sites within 1Kb upstream of the transcription start. Mean and standard deviation are given to 1 d.p.

If these patterns of bases were occurring in the upstream regions of the ORFs known to be essential for stationary phase at random the distribution of matches would be expected to follow a Poisson distribution. The results of the genome-wide pattern matching can be used to calculate the expected (Poisson) distribution of the occurrence of binding sites using the formula:

$$\text{Expected number of genes with "x" number of matches} = \left(\frac{e^{-\lambda}.\lambda^{x}}{x!}\right).N$$

Where λ is the average number of matches per upstream sequence for the entire genome and N is the total number of genes (table 4-6). A χ^2-test can then be used to determine whether the observed distribution of matches is significantly different from the expected Poisson distribution:

$$\chi^2_{\alpha} = \Sigma\left(\frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}\right)$$

For this calculation the observed/expected numbers and the null hypothesis are:

Observed the number of stationary phase essential ORFs with "n" matches to a pattern.

Expected the expected number of stationary phase essential ORFs with "n" matches to a pattern.

Null Hypothesis (H_0) "*The matches to each pattern in the region 1Kb upstream of stationary phase essential ORFs occur at random and thuis follow a Poisson distribution*"

The χ^2-test was performed using the CHITEST function on Microsoft Excel. If the probability returned by this function is less than 0.05 the null hypothesis will be rejected. The results of the test are shown in table 4-7.

Number of sites		Binding Sites				Upstream Activation Sequences			
		Adr1p	Gln3p	Mig1p	Swi5p	CRE	CSRE	*PDS*	STRE
Observed	0	42	49	87	38	91	92	91	68
	1	35	37	4	35	1	0	1	21
	2	12	5	0	14	0	0	0	3
	3	3	0	1	3	0	0	0	0
	4	0	0	0	2	0	0	0	0
	5	0	1	0	0	0	0	0	0
	6	0	0	0	0	0	0	0	0
Expected	0	39.3	56.1	86.7	45.4	91.7	92.0	91.1	60.5
	1	33.4	27.7	5.1	32.1	0.3	0.0	0.9	25.3
	2	14.2	6.9	0.2	11.3	0.0	0.0	0.0	5.3
	3	4.0	1.1	0.0	2.7	0.0	0.0	0.0	0.7
	4	0.9	0.1	0.0	0.5	0.0	0.0	0.0	0.1
	5	0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.0
	6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 4-6: The observed and expected distributions of the number of binding sites upstream of the 92 SGDP positive genes (expected figures are given to 1 d.p.).

Site	*Binding Sites*				*Upstream Activation Sequences*			
	Adr1p	*Gln3p*	*Mig1p*	*Swi5p*	*CRE*	*CSRE*	*PDS*	*STRE*
χ^2 *TEST*	*1.0*	3×10^{-12}	7×10^{-66}	*0.7*	*1.0*	*1*	*1*	*1.0*

Table 4-7: Result of the χ^2-test on the observed & expected figures (to 1 d.p.).

Most of the sites do not appear to have a distribution of binding sites for the YPD92 set of ORFs that is significantly different from the occurrence of binding sites across the whole genome. The binding sites for Mig1p & Gln3p however do seem to be different. If

these results are examined more closely it can be seen that, in each distribution, there is one occurrence of an upstream region with more than the usual number of matches in it. For the Mig1p binding site there is one ORF that has three potential binding sites within 1Kb of its transcription start point while for the GLn3p binding site there is one ORF that has five potential binding sites within 1Kb of its transcription start point. If the χ^2-test is repeated on these two distributions, but with the possibly anomalous results ignored (changed to a value of zero), the conclusion is changed to the observe distribution is not different from the expected (Poisson) distribution (table 4-8). It is possible that that the non-Poisson distribution of these two patterns indicates that they do not occur at random. However, in both cases, this is because of one occurrence of more matches than is expected and the removal of these occurrences drastically alters the result of the χ^2-test so they may well be anomalous results and no conclusion is drawn.

Number of hits	**Gln3p binding site**		**Mig1p binding site**	
	Observed	**Expected**	**Observed**	**Expected**
0	49	56.1	87	86.7
1	37	27.7	4	5.1
2	5	6.9	0	0.2
3	0	1.1	1	0.0
4	0	0.1	0	0.0
5	1	0.0	0	0.0
6	0	0.0	0	0.0
7	0	0.0	0	0.0
8	0	0.0	0	0.0
9	0	0.0	0	0.0
10	0	0.0	0	0.0
χ^2 TEST	2.81×10^{-12}		6.53×10^{-66}	
χ^2 TEST; ignoring the potentially anomalous result	0.8		1.0	

Table 4-8: Potentially anomalous results are highlighted in grey. Figures given to 1d.p.

SPEX program

SPEX (at http://ep.ebi.ac.uk) was used to test the gene sequences of the stationary phase essential genes to look for sequences over-represented in this set of genes. The SPEX program as run using the YPD105, YPD92, and RANDOM100 sets of upstream sequences, with default settings with the exception of:

Background sequences: YEAST_-1000_+2_W_all.fa

Return patterns with a probability of less than: 1

The SPEX program is trying to locate sequences that are over represented in the 1 Kb upstream sequences of the genes of interest when compared to the rest of the genome. The YEAST_-1000_+2_W_all.fa file was used for the background sequences because it contains the 1 Kb upstream sequences of the entire *S. cerevisiae* genome. The return probability was set to 1 so the program would return any sequence for which less than one occurrence is expected.

The SPEX program found no patterns in the 105 upstream sequences of the gene knockouts identified in the first screen. Once again no patterns were discovered when only upstream sequences of the 92 knockout strains that were identified as having a stationary phase defect (in YPD medium) in the secondary screens suggesting that there were no sequences that were over represented in the upstream sequences of the stationary phase essential genes. The result of no identified patterns was repeated when the control set of 100 random genes was used.

Pattern discovery

The Oligo Analysis program (http://rsat.ulb.ac.be/rsat/) was used to locate over represented patterns in the YPD92 set of sequences. The program was set to examine the sequences in the YPD92 set and not their complementary sequences. The pre-defined background frequencies were set as: "upstream-noorf *Saccharomyces cerevisiae*" as the sequences being used are intergenic sequences upstream of the genes of interest although

there will be overlap with the coding sequences of the adjacent upstream ORFs in many, if not all, of the sequences. The program was run three times, looking for patterns of length 6, 7, and 8 bp. The result of these searches is shown in Table 4-9.

A total of sixteen sequences of lengths 6 bp, eight 7 bp, and four 8 bp were identified. Many of these patterns could be part of a larger consensus sequence (Vilo, Brazma et al. 2000). Identified patterns were collated and then aligned using the Magi program (http://www.hgmp.mrc.ac.uk). The default settings were used, apart from output order being set to aligned instead of input. The alignment of sequences by the Magi program is shown in Diagram 4-3 (left column).

As the patterns have been collected by three different operations of the same program using the same set of genes looking for different length patterns, some of the patterns have been identified multiple times. For example the patterns oligo6-6 and oligo7-6 would appear to be the same pattern. The multiple sequence alignment aligned all the patterns into one group because the program tries to align all the sequences it is given together. After a visual examination the aligned sequences were split into 4 different groups (Diagram 4-3, right column) as individually each of these groups appeared to be much more closely related. In addition, a search of the upstream sequences was for any over-represented sequences not just one particular sequence. It is therefore quite likely that more than one sequence was found. For each of the groups a consensus sequence was constructed by examining each column of aligned bases and assigning a consensus base at each position according to the predominant base(s) in the column.

pattern	Occurrences		Occurrence (binomial)			rank
	observed	expected	probability	E-values	significance	
acaatg	59	26.15	2.4e-08	9.7e-05	4.01	1
catttg	59	31.61	8.6e-06	3.5e-02	1.45	2
ggtttg	40	18.80	1.4e-05	5.8e-02	1.24	3
tttgga	58	31.81	1.9e-05	7.9e-02	1.10	4
gctggt	28	11.82	4.4e-05	1.8e-01	0.75	5
cagatg	32	14.63	5.7e-05	2.4e-01	0.63	6
ggtcaa	30	13.47	7.1e-05	2.9e-01	0.54	7
tggaac	34	16.33	8.8e-05	3.6e-01	0.44	8
tggcaa	47	25.58	9.3e-05	3.8e-01	0.42	9
accatt	52	29.43	0.00011	4.4e-01	0.36	10
tccaat	46	25.09	0.00012	4.7e-01	0.33	11
ctggtg	27	12.13	0.00016	6.4e-01	0.19	12
ccaatg	32	15.59	0.00018	7.3e-01	0.14	13
tggacg	20	7.81	0.00019	7.6e-01	0.12	14
gcctgt	26	11.69	0.00021	8.5e-01	0.07	15
taccat	46	25.81	0.00021	8.7e-01	0.06	16
ataccat	24	8.27	6.40E-06	1.00E-01	0.98	1
ctcagca	14	3.23	7.70E-06	1.30E-01	0.9	2
gctggtg	13	2.84	9.30E-06	1.50E-01	0.82	3
atcccag	12	2.48	1.20E-05	1.90E-01	0.71	4
ctgacaa	17	5.27	3.80E-05	6.20E-01	0.21	5
gacaatg	16	4.77	4.00E-05	6.60E-01	0.18	6
gtccaat	14	3.81	4.60E-05	7.50E-01	0.13	7
tcagcaa	20	7.09	5.20E-05	8.60E-01	0.07	8
ctcagcaa	9	0.9	4.70E-07	3.10E-02	1.51	1
atcccaga	8	0.79	1.80E-06	1.20E-01	0.92	2
cataccat	10	1.53	4.90E-06	3.20E-01	0.49	3
ttgcaaag	12	2.33	6.40E-06	4.20E-01	0.38	4

Table 4-9: Results of the Oligo Analysis program

	oligo6-1 ---ACAATG---
	oligo7-5 --GACAATG---
oligo6-1 ---ACAATG---	oligo7-7 -GTCCAAT----
oligo7-5 --GACAATG---	oligo6-12 --TCCAAT----
oligo7-7 -GTCCAAT----	oligo6-5 ---CCAATG---
oligo6-12 --TCCAAT----	oligo7-1 -ATACCAT----
oligo6-5 ---CCAATG---	oligo8-2 CATACCAT----
oligo7-1 -ATACCAT----	oligo6-11 --TACCAT----
oligo8-2 CATACCAT----	oligo6-2 ---ACCATT---
oligo6-11 --TACCAT----	oligo7-2 -ATCCCAG----
oligo6-2 ---ACCATT---	oligo8-1 -ATCCCAGA---
oligo7-2 -ATCCCAG----	**CONSENSUS CATMCMATG**
oligo8-1 -ATCCCAGA---	oligo7-3 --CTCAGCA---
oligo7-3 --CTCAGCA---	oligo8-3 --CTCAGCAA--
oligo8-3 --CTCAGCAA--	oligo7-8 ---TCAGCAA--
oligo7-8 ---TCAGCAA--	oligo7-4 CTGACAA-----
oligo7-4 CTGACAA-----	oligo6-15 -TGGCAA-----
oligo6-15 -TGGCAA-----	oligo6-9 -GGTCAA-----
oligo6-9 -GGTCAA-----	**CONSENSUS CTCRGCAA**
oligo6-8 --GCTGGT----	oligo6-8 --GCTGGT----
oligo7-6 --GCTGGTG---	oligo7-6 --GCTGGTG---
oligo6-6 ---CTGGTG---	oligo6-6 ---CTGGTG---
oligo6-3 ---CAGATG---	oligo6-3 ---CAGATG---
oligo6-7 --GCCTGT----	**CONSENSUS GCTGGTG**
oligo6-4 -CATTTG-----	oligo6-7 --GCCTGT----
oligo6-10 -GGTTTG-----	oligo6-4 -CATTTG-----
oligo6-16 ---TTTGGA---	oligo6-10 -GGTTTG-----
oligo8-4 ----TTGCAAAG	oligo6-16 ---TTTGGA---
oligo6-13 -----TGGAAC-	oligo8-4 ----TTGCAAAG
oligo6-14 -----TGGACG-	oligo6-13 -----TGGAAC-
	oligo6-14 -----TGGACG
	CONSENSUS GTTTGGAA

Diagram 4-3: Identified sequences aligned using the MAGI program.

The four consensus sequences could potentially be part of a regulatory sequence important for the set of genes that were used in this investigation. The *Saccharomyces cerevisiae* Promoter Database (SCPD) (http://cgsigma.cshl.org/jian/) is a database for regulatory sequences in yeast. By searching the database with possible regulatory sequences it is possible to identify those sequences that could be part of a known regulatory element. This was done with the four consensus patterns discovered using the Oligo Analysis program. All regulatory factor consensus sequences come from SCPD, the UASH consensus sequence from Gailus-Durner *et al.*, and the GFI consensus sequence from Dorsman *et al.* (Dorsman, Doorenbosch et al. 1989; Gailus-Durner, Xie et al. 1996).

The sequence CATMCMATG did not exactly match any known pattern in the SCPD. No matches were also reported when up to one mismatch was allowed. When up to two mismatches were allowed there were many matches. However these were not considered because allowing two mismatches on a sequence of only 9 bp allows too much variation from the consensus sequence. Another piece of evidence suggesting that this is a real sequence is that the sequence appears to be a palindrome. The complementary sequence of this pattern is CATKGKATG. If the sequence is palindromic the central three bases are not following the pattern that they should. This could be because they are not bases at which there is a base requirement. In which case the consensus sequence would be $CATN_3ATG$.

There are not any exact matches in the SCPD to the sequence CTCAGCAA. When one mismatch was allowed to occur two matches were found: The UASH site from the *RFA1* gene and the *BAS1* site from *ADE2*. However these matches were to the complement of the pattern (TTGCTGAG).

UASH (consensus: TNTGNWGT): CATTTTGCTGTGCTG

BAS1 (consensus: TAAWTK): AGTGATTGACTCTTGCTGACCT

The complement of this pattern seems to match the UASH consensus sequence. While it also matches part of the *BAS1* site form the *ADE2* promoter it does not overlap the

consensus sequence for *BAS1* and so probably does not indicate that this sequence is anything to do with the *BAS1* promoter.

The sequence GTTTGGAA did not show homology with any known pattern in the SCPD. When one mismatch was allowed to occur three matches to the complement of the pattern was found. Two matches (each with one mismatch) with the *DAL82* site in the gene *DAL7*. The third complementary match of the pattern is to the *GFI/TAF* site in the *QCR2* gene.

DAL82 (No consensus):

GGTGCGATAGAATACCGCGGATTTTGGAAAATTGCGTTTGCTTTTCTTATCACATA

GCTGAAAGTTGCGGTGCGATAGAATACCGCGGATTTTGGAA

GFI/TAF (consensus: RTCRYN$_5$ACG): CTGATCATTCCCAACGAACCAATAG

The complement of this pattern seems to match the *GFI/TAF* consensus sequence. However a lot of the match falls into a region of the *GFI/TAF* consensus sequence with no strict base requirement, making this a weaker match than it first appears.

The sequence GCTGGTG did not exactly match any known pattern in the SCPD. When one mismatch was allowed to occur a several matches were found. There were matches to the *ABF1* site in the *PGK1* gene, *HAP2/HAP4/HAP4* site in *CYC1*, and the *PHO4* site in *PHO84*. The complement of the pattern also matched the *PHO2/SWI5* and *SWI5* sites in the *HO* gene, the *RAP1* site in *ENO1*, the *GAL4* site in *GAL2*, and the *ABF1* site in *RPO31*.

ABF1 (consensus: TCRN$_6$ACG): GAAATTACCGTCGCTCGTGATTTGTTTG

HAP2/HAP3/HAP4: CGTTGGTTGGTGGA

PHO4 (consensus: CACGTK): TTACGCACGTTGGTGCTG

PHO2 (consensus: TAAWTK): TTTAAAAAAAAAACCAGC

SWI5 (consensus: KGCTGR): AAAAACCAGCATGCTATAATGCT

RAP1 (consensus: RMACCCA): GCACCCAAACACCTGCATATTTGG

GAL4 (consensus: CGGN$_{11}$CCG): CACCGGCGGTCTTTCGTCCGTGC

ABF1 (consensus: $TCRN_6ACG$): ATCACCAGATCAAATAACAGAA

The only promoter sequence with which there is significant overlap with this sequence is the *ABF1* consensus sequence. However most of this overlap is to the central portion of the consensus sequence where there is no sequence requirement. Thus this is probably not a match.

The four patterns identified all represent possible regulatory sequences used in the regulation of the genes identified to be essential for the maintenance of viability in stationary phase. The pattern CATMCMATG does not seem to match any known promoter sequence. Thus could be a binding site for an as yet unidentified transcription regulatory protein that is essential for stationary phase. That the sequence appears to be palindromic adds strength to this conclusion.

GCTGGTG could also be the site for a novel stationary phase transcription factor. This pattern matches on several occasions to *ABF1* and *RAP1* consensus sequences. Rap1p & Abf1p are global regulators of ribosome biosynthesis (Planta, Goncalves et al. 1995; Gailus-Durner, Xie et al. 1996). The set of genes used in this investigation had a considerable number of ribosomal encoding genes (approximately 20% of the genes identified). Thus this result could be due to the set of sequences from which this sequence was extracted.

The complement of the sequences CTCAGCAA and GTTTGGAA seem to match the UASH and *GFI/TAF* sites. GFI binds to the consensus sequence $RTCRYN_5ACG$ and is probably the same protein as the proteins SUF and TAF (Dorsman, Doorenbosch et al. 1989). These proteins are trans-activators of the constitutive transcription of the genes for ribosomal proteins S33 and L3 respectively suggesting that this pattern could have been identified due to the ribosomal bias of the set of genes used. There is no obvious reason why UASH sites should be over represented in the set of genes used in this study. UASH sites are often found in meiosis specific genes, and in the upstream region of *HOP1* it has

been shown that Abf1p (autonomously replicating sequence-binding factor 1) can bind to UASH (Gailus-Durner, Xie et al. 1996; Mercier, Denis et al. 2001).

Sequence Analysis Program

The four consensus patterns were put into the sequence analysis program. This program finds all the genes with the input sequence in their upstream region. Using results of transcript analysis experiments it then identifies those genes that are co-ordinately regulated (Ihmels, Friedlander et al. 2002). When this program was used to analyse the four patterns identified it revealed that all four patterns do not appear in any co-ordinately regulated set of genes. Examination of the recurrence profiles returned by the program of the four sets defined by the input sequences shows that they act similar to random gene datasets (Diagram 4-4). This suggests that these sequences are not regulatory sequences.

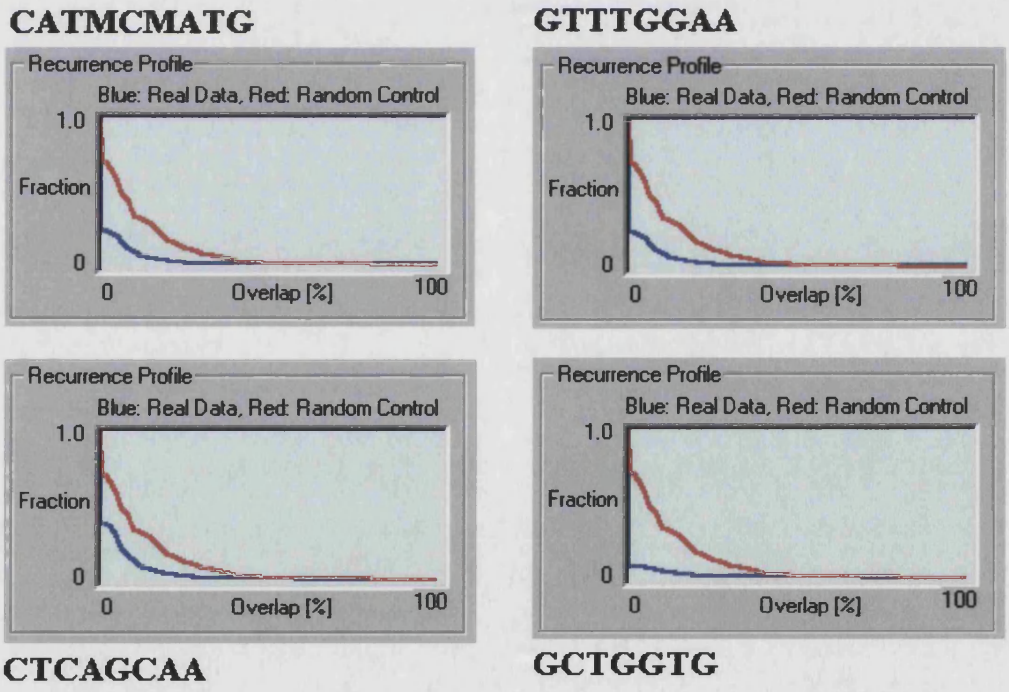


Diagram 4-4: Recurrence Profiles. This plots the percentage overlap between the transcription profile of the submitted set of genes with the profile of a set made of the submitted genes with a number of random genes versus the fraction of the second set of genes that is made from random genes.

Discussion

The clustering of genes by their transcript profiles revealed little that in itself was novel and/or unexpected. When the transcript profiles were ordered using the program Cluster it is possible to observe genes of unknown function grouped among many genes of a similar or same function. This information is not very useful on its own. However when combined with other information it can be used to help confirm and strengthen the case for other suppositions. For example, the ORF YNL177c is grouped alongside genes that are known to be involved in protein synthesis. While it has yet to have a function assigned to it has similarity to *Thermotoga maritima* ribosomal protein L22 and *Schizosaccharomyces pombe* mrpl22 (mitochondrial ribosomal protein subunit L22 - information from the Yeast Protein Database annotation). These two pieces of information taken together provide a much stronger argument as to the function of the protein coded for by the ORF YNL177c being mitochondrial protein synthesis.

In a similar way to the results of the gene ordering by the program Cluster the results of the Pathway Analysis Program are not, when taken in isolation, particularly revealing about the function or relationships of those ORFs identified as being essential for stationary phase. The results of the program suggested that the inputted genes (those identified as being essential for stationary phase) are part of a larger mitochondrially related set of genes. This fits in well with other information obtained in this investigation that also points to this conclusion (such as respiration, which occurs in the mitochondrion, being important for stationary phase survival) and helps to further support these conclusions.

Four potential stationary phase regulatory sequences were identified by analysing the upstream sequences of genes identified as being essential for stationary phase. This is however a prediction and is not based on any physical evidence. To investigate whether these sequences are stationary phase regulatory elements band-shift and promoter mutation assays should be conducted. If the sequences identified are stationary phase regulatory

sequences mutation/deletion should result in the same phenotype as a deletion of the gene that they are regulating. Oligonucleotides that contain each of the sequences could be constructed and then incubated with no cell extract, or cell extract from exponential, post-diauxic and stationary phase cells. If the sequences are required for stationary phase the movement of the oligos through a gel should be affected when they have been incubated with the post-diauxic and/or stationary phase cell extracts. Stationary phase regulation may be seen in the post-diauxic phase as well as stationary phase because of the change in the levels of transcription and translation upon entry to stationary phase. Transcription and translation in stationary phase is very low, and no novel stationary phase specific proteins have been identified. This means that any major changes required for stationary phase must occur before entry into stationary phase, that is, during the post-diauxic phase. If a band-shift is observed it may also be possible to purify and identify the protein responsible for this.

Chapter 5

Potential IRES sites in *S. cerevisiae*

Summary

Global transcription rate is down-regulated in cells of the yeast *Saccharomyces cerevisiae* on entry into stationary phase. It has been proposed that genes specifically required at that time can be more highly expressed by recruiting ribosomes at Internal Ribosome Entry Sites (or IRES). We performed a large scale screen of individual gene knockouts for the entire genome for the ability to maintain long-term viability in stationary phase. 102 of the ORF knockouts tested showed a significant loss in viability. We have examined the upstream regulatory regions of our candidate genes for potential IRES sites. We found no evidence for this subset having a significantly different number of such potential sites compared to control data sets.

Internal Ribosome Entry Sites in *S. cerevisiae*

Translation of eukaryotic mRNAs occurs mainly using the ribosome scanning mechanism in which a ribosome attaches to the 5'-CAP structure of the mRNA and scans along the mRNA until it reaches the start codon of the open reading frame (ORF) (reviewed in (McCarthy 1998; Kozak 1999; Sachs and Varani 2000). Internal transcription initiation has also been observed when a ribosome is recruited to an Internal Ribosome Entry Site (IRES) independently of the 5'-CAP structure. IRES have been shown to be active in viruses and mammalian cells and have recently been demonstrated to be present and active in starved *S. cerevisiae* cells (Paz, Abramovitz et al. 1999; Paz and Choder 2001; Zhou, Edelman et al. 2001). For those IRES sequences that have been characterised there is a great diversity in primary sequence and secondary structure. This has prevented the delineation of a consensus sequence and the construction of a detailed model on how

IRES sequences might work (Martínez-Salas, Ramos et al. 2001; Pestova, Kolupaeva et al. 2001). Recent work on crucifer-infecting tobamovirus (crTMV) showed that polypurine (A)-rich sequences are responsible for IRES activity in this virus. These elements were also shown to be active in plant, HeLa cells and yeast (Dorokhov, Skulachev et al. 2002).

A number of laboratories (Tranque, Hu et al. 1998; Hu, Tranque et al. 1999; Zhou, Edelman et al. 2001) have shown that homology of the 5'-UTR of the mRNA with the 3' end of the 18S rRNA may be significant in internal translation initiation by complementary base pairing. Initiation factor eIF4E interacts with the 5'-CAP and eIF4E-independent translation has been shown to be necessary and sufficient for cell survival in stationary phase (Paz and Choder 2001). It has been suggested that, by having a novel means of translation initiation, genes with IRES can be more highly expressed during stationary phase at times when global transcription is down-regulated to about 1% of the rate of exponentially growing cells (Paz and Choder 2001).

A complete gene knockout set of the yeast *S. cerevisiae* was created by the Saccharomyces Genome Deletion Project (http://www-sequence.stanford.edu/group/yeast_deletion_project/deletions3.html). We have previously done a complete screening for the ability of knockout colonies to remain viable in long-term stationary phase after starvation. Of the 5883 ORF knockouts tested, 102 showed at least a 10^6-fold loss of viability over 90 days at 28°C. These genes were therefore classified as being essential for stationary phase viability. We hypothesised that such genes might be good candidates for genes with IRES. Here we investigated whether the 5'-UTR of this subset of genes contained sites that could allow base pairing of the mRNA with the 18S rRNA in the ribosome and thus help initiate transcription via internal ribosome entry.

Materials and Methods

Databases

Sequence Retrieval System (SRS; http://srs.hgmp.mrc.ac.uk/SRS6/) was used to retrieve all known complete *S. cerevisiae* 5'-UTRs from the EMBL database. A total of 531 *S. cerevisiae* complete 5'-UTR sequences were obtained from this (5UTR file) and have an average size of 146 bp.

Sequences 1Kb upstream of all ORFs (where the first base is the base adjacent to the A in the ATG start codon) were downloaded using the GENOMES tool from the Expression Profiler website (http://ep.ebi.ac.uk/EP/) to create the Upstream file. These sequences represent upstream regions of genes for the entire genome but may contain sections of coding sequence where upstream genes are close together.

A multiple sequence file (intergenic file) containing genomic intergenic sequences between ORF and RNA genes, LTR and Ty sequences of *S. cerevisiae* was downloaded from the *Saccharomyces* Genome Database (ftp://genome-ftp.stanford.edu/yeast/yeast_NotFeature). Like the sequences in the upstream dataset, these sequences represent 5'-UTRs of genes but will not contain any coding sequences. The 6321 sequences in this dataset have an average length of 484 bp.

A set of randomly generated sequences (for use as a control) was generated using the random sequence generator from the RSA-tools website (http://bio.cigb.edu.cu/jvanheld/rsa-tools/random-seq.html). Six thousand sequences were generated, each 1 Kb long and with a base composition equivalent to *S. cerevisiae* non-coding regions.

Sequence comparisons

The three principal sets of upstream sequences were used to set up a BLAST database for use with the WU-BLAST program using the pressdb program supplied with WU-BLAST (Table 5-1). The sequences of the 25S (Accession Number: SCRGIH5), 18S (Accession Number: SCZ75578), 5.8S (Accession Number: SC09327), and 5S (Accession

Number: SCRNA9). rRNA genes were then queried against these databases using WU-BLAST. The 18S rRNA sequence was used to search the sequence databases as base pairing between upstream sequences and this rRNA have been suggested to be instrumental in the initiation of internal transcription initiation (Tranque, Hu et al. 1998; Hu, Tranque et al. 1999; Zhou, Edelman et al. 2001). As a control, sequences for the other three rRNAs were also used. If base pairing with the 18S rRNA is the method of IRES transcription initiation there should be similarity between the upstream sequences and the 18S rRNA but not the upstream sequences and any of the other rRNAs.

WU-BLAST command line/search settings were (where italicised words indicate a variable entry): blastn *databasefile querysequence* E=10000 W=7 H=0 V=0 B=2000000000 –altscore "T C 5" –altscore "G A 5" >*output*.txt. The values of E and W restrict the BLAST hits to a maximum E-value of 10000 with at least one region with 100% identity over 7 or more bases in length. The value of B means that up to 2×10^9 hits will be reported. The two *–altscore* arguments allow for GU mismatching (appendix 5) (Varani and McClain 2000). Selected information from the results of the BLAST search (name of sequence, length of hit, score, E-value, probability, identity, positives, length of hit, strand query start and subject start for each BLAST hit) were then extracted from the output files of the BLAST search using a perlscript program (appendix 6). From the results of the Upstream dataset a subset was created called "SPD" file that consisted of only those genes for which a stationary phase essential phenotype has been determined. A summary is given in Table 5-1.

Name of BLAST database	Number of sequences in file	Description of sequences
5UTR	531	Known 5'-UTR sequences, downloaded from the EMBL database.
Intergenic	6321	Intergenic sequences from (without mitochondrial sequences) downloaded from the SGD website.
Upstream	6423	Upstream sequence (1000bp) of all ORFs.
Random	6000	Randomly generated sequences with a base composition equivalent to *S. cerevisiae* non-coding regions.
SPD	105	Sequences from the Upstream set, whose gene knockouts show a stationary phase lethal phenotype

Table 5-1: Contents of the BLAST databases set for use in this study.

Poisson analysis

χ^2 can be calculated using the formula:

$$\chi^2 = \Sigma \frac{(O-E)^2}{E}$$

Where O and E are the observed and expected values respectively. The expected values can be calculated from the average number of hits per sequence for each set of genes and the total number of sequences in each set using the formula:

$$\text{Expected number of sequences with "N" BLAST hits} = \text{Total number of sequences} \times \frac{e^{-\lambda}\lambda^N}{N!}$$

Where X is the number of hits per database sequence, and λ is the average number of BLAST hits per database sequence. The null hypothesis for the χ^2 test is that there is no difference between the observed and calculated values. A test for normality was calculated using the Kolomogorov-Smirnov normality test function on Minitab version 12.

Results

Occurrence of BLAST hits

If the 18S rRNA is the rRNA molecule with which base pairing to mRNA occurs (as has been proposed; (Hu, Tranque et al. 1999), then the 18S rRNA should be more likely to have regions of homology to mRNA 5'-UTR sequences than with other sequences that are not 5'-UTRs. These regions that allow base pairing should not be seen when 5'-UTRs and the 25S, 5.8S and 5S rRNAs are compared. Furthermore, if IRES are active in stationary phase in *S. cerevisiae*, then the number of regions of homology is likely to be greater when only stationary phase essential genes are examined.

If a sequence occurs in an rRNA gene with homology to the 5'UTR of another gene, the sequence in the 5'-UTR of that gene may be in one of two orientations; it can be read in the forward or reverse directions (Diagram 5-1). In the "reverse orientation" the motif of interest can be read (in the 5' to 3' direction) on the non-coding strand of the DNA. This, when transcribed, produces a sequence in the mRNA which could potentially base pair with rRNA. When the motif is read on the coding strand ("forward orientation") a sequence is produced in the mRNA that exactly matches the motif in the rRNA. This could allow the mRNA to bind the ribosomal protein in place of the rRNA. For base pairing to be able to occur, a sequence in the 5'-UTR of a gene must be complementary to a sequence occurring in the rRNA gene. Using this model, sequence homology (in the "reverse orientation") was compared to the forward orientation, which is the biologically significant orientation.

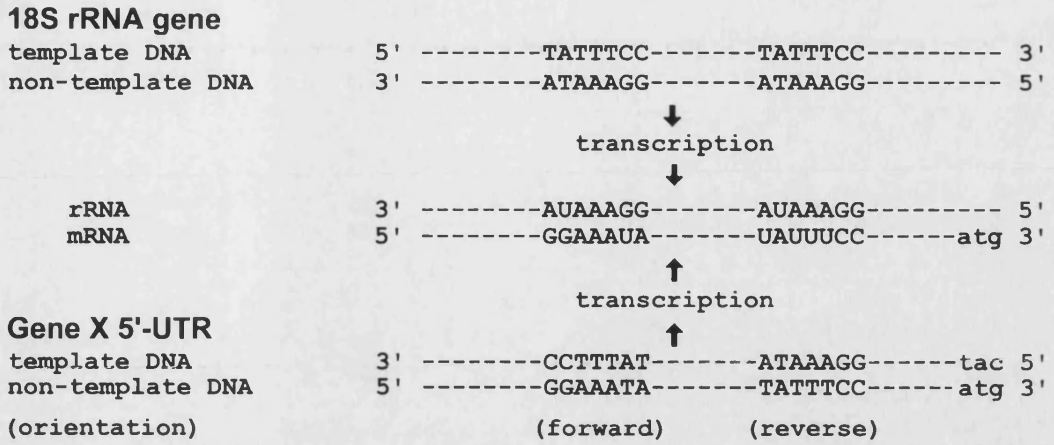


Diagram 5-1: Complementary pairing between mRNA and rRNA. Only one of the two possible orientations of a homologous sequence allows base pairing between the mRNA and the rRNA to occur.

In most cases the number of matches relates to the reverse order of rRNA gene sequence length (Diagram 5-2). When comparing datasets the 18S rRNA does not generate a disproportionate number of hits, and the SPD set does not have a higher hit rate compared to the other datasets. The results indicate that if there are IRES sites they do not occur more frequently in the SPD set and/or that their frequency must be low since they do not appear at a significantly higher frequency than by random expectations.

Nature and distribution of the BLAST hits

Since the previous results show that the number of reverse orientation matches is inconsistent with the hypothesised role in stationary phase, what is the significance of the hits? The Null hypothesis is that the matches are random and of no biological significance. Accordingly, the frequency distribution on individual UTRs (Diagram 5-3) should follow a Poisson distribution as an event occurring at low frequency in space or time. This prediction can be tested with a χ^2 test using the average number of BLAST hits per sequence to generate an expected Poisson distribution. The P-values of χ^2 test were

calculated and are shown in Table 5-2. A Kolmogorov-Smirnov Normality test was completed for the same set of distributions using Minitab (Table 5-3).

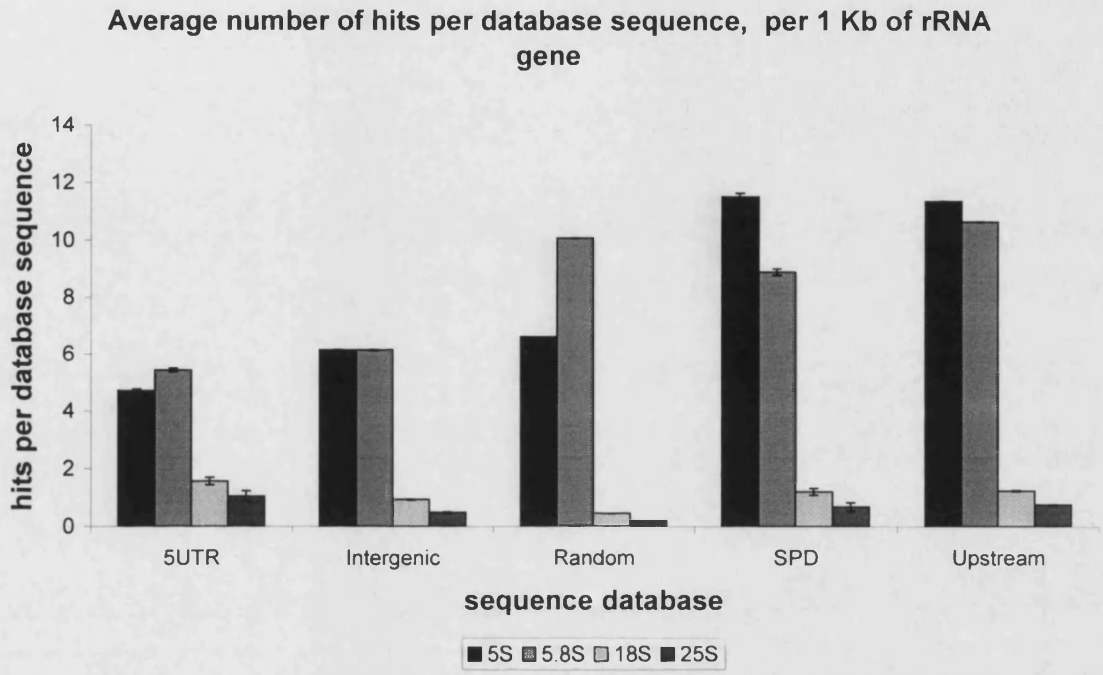


Diagram 5-2: BLAST hit frequency. The average number of BLAST hits (in the reverse orientation only) for each database sequence per 100bp of rRNA gene sequence (5S – black; 5.8S – black lines/white background; 18S – grey; 25S – white lines/black background), when using BLAST to compare the *S. cerevisiae* rRNA gene sequences against databases of real and synthetic 5'-UTR. The error bars represent the standard error of the mean for each column.

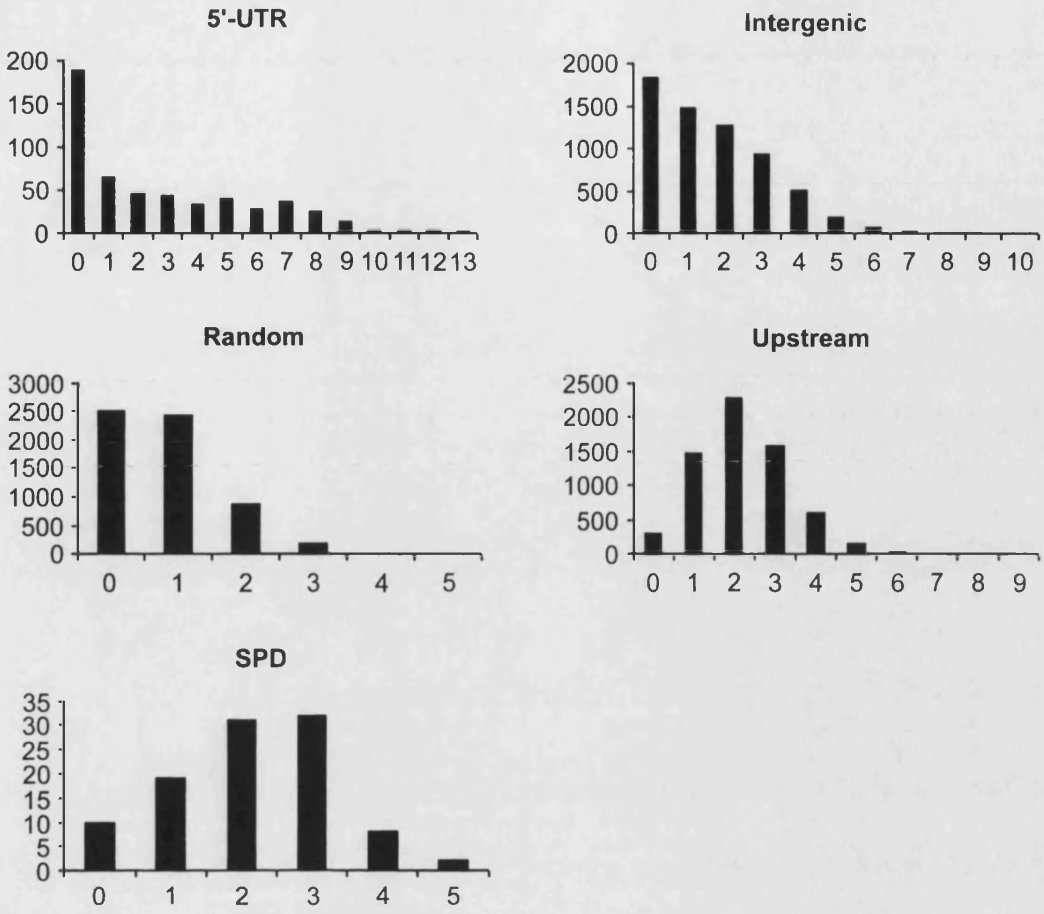


Diagram 5-3: BLAST hit frequency. Distribution (relative frequencies, Y-axis) of the number of BLAST hits per database sequence (X-axis) for each of the sets of genes with the 18S rRNA.

rRNA	*p* values (χ^2 test), reverse orientation				
	5UTR	**Intergenic**	**Random**	**SPD**	**Upstream**
25S	< 0.01	< 0.01	0.036	> 0.15	< 0.01
18S	< 0.01	< 0.01	< 0.01	0.032	< 0.01
5.8S	< 0.01	< 0.01	< 0.01	> 0.15	< 0.01
5S	< 0.01	< 0.01	> 0.15	> 0.15	> 0.15

Table 5-2: P-values for the χ^2 test for a Poisson distribution (3 d.p.).

rRNA	*p* values (normality test), reverse orientation				
	5UTR	**Intergenic**	**Random**	**SPD**	**Upstream**
25S	< 0.01	< 0.01	< 0.01	> 0.15	< 0.01
18S	< 0.01	< 0.01	< 0.01	> 0.15	< 0.01
5.8S	< 0.01	< 0.01	< 0.01	> 0.15	< 0.01
5S	< 0.01	< 0.01	< 0.01	> 0.15	< 0.01

Table 5-3: Kolmogorov-Smirnov normality test.

Unlike the other four datasets, the blast hits in the SPD set seem to be normally distributed. This is almost mirrored in the χ^2 test used to determine if the distributions follow a Poisson distribution. The SPD distributions seem to follow a Poisson distribution, with the exception of the SPD/18S distribution (at the 5% significance level). The Random/5S and Upstream/5S distributions also seem to follow a Poisson distribution.

These results suggest that there is something different about the SPD set of genes, when compared to all or a random selection of genes/sequences. Examining the average number of hits per sequence shows that the arithmetic mean is no larger than any other distribution. However the median and mode suggests that this set has slightly more hits per database sequence than the other sets (Table 5-4). The lack of any obvious trend or difference with the SPD set of genes suggests that whatever is causing the difference between this set of genes and the other sets, it is small. The median and mode data suggest that not all of the SPD genes have this difference which is why there is no obvious difference between the SPD set and the other sets.

Set	Mean	Mode	Median	St. deviation
5UTR	2.82	0	2	3.02
Intergenic	1.65	0	1	1.53
Random	0.80	0	1	0.82
Upstream	2.21	2	2	1.16
SPD	2.15	3	2	1.17

Table 5-4: The average number (and standard deviation) of hits per sequence for the 18S/reverse orientation distributions.

Location of the BLAST hits on the rRNA

In prokaryotes translation initiation occurs by using base pairing between the mRNA and the prokaryotic equivalent of the 18S rRNA using the Shine-Dalgarno sequence. This sequence allows base pairing between the mRNA and the 3' end of the rRNA. If the same process is occurring for IRES in S. *cerevisiae* a similar pattern might emerge. However similar studies of possible IRES sites in other eukaryotic systems have found regions in the rRNA equivalent to the *S. cerevisiae* 18S rRNA that could possibly be used to base pair with 5'-UTRs throughout its length, not just at the 3' end of the rRNA (Mauro and Edelman 1997; Tranque, Hu et al. 1998).

A plot along the UTR of the start point for each BLAST was made (Diagram 5-4). This shows a relatively similar number of BLAST hits along the whole length of the 18S rRNA, with a peak of BLAST hits at either end of the sequence. If the BLAST hit start points for each of the orientations is examined separately (Diagram 5-4) it can be seen that the peaks at each end are comprised solely of a BLAST hit start point from one orientation. The peak at the 5'-terminus of the 18S rRNA sequence is comprised of only hits from the forward orientation, while the peak at the 3'-terminus of the 18S rRNA sequence is comprised of only hits in the reverse orientation.

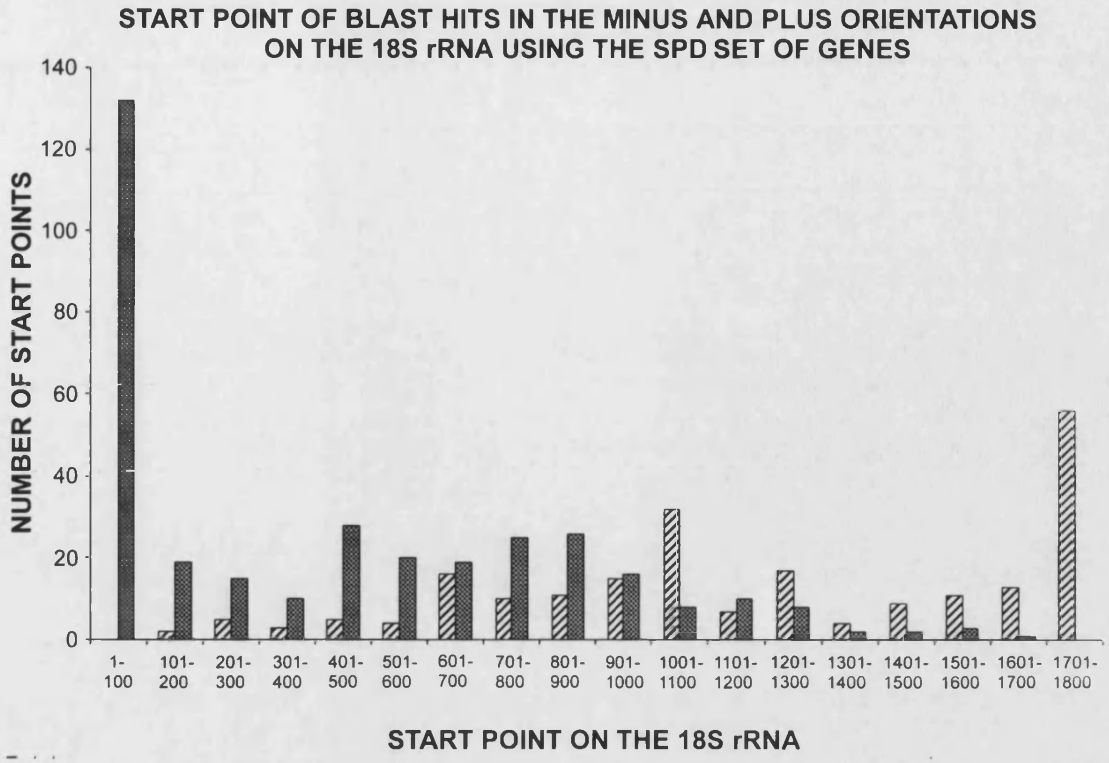


Diagram 5-4: Location of BLAST hits. The start points of BLAST hits on the 18S rRNA, from the "SPD" set of genes (reverse orientation – black lines/white background; plus orientation – white crosshatching/black background).

This result suggests that the BLAST search has found some evidence of IRES. If IRES function in a manner analogous to prokaryotic translation initiation there should be a peak of BLAST start points (in the reverse orientation) at the 3'-terminus, which can be seen (Diagram 5-4). However the pattern of BLAST hit start points for the forward orientation mirrors the pattern formed by the hits in the reverse orientation with a peak at the 5'-terminus. Also this pattern of peaks of start points at either end of the rRNA sequence is seen for all sets of genes (including the randomly generated sequences) with all four rRNA sequences suggesting that it is an artefact of the search process and does not have a biological basis. Reverse orientation hits would not be expected to be found at the 5'–terminus of the rRNA sequence as there would not be room for the region of homology to extend before the end of the rRNA sequence. The same argument can be used for plus

orientation BLAST hits at the 3'-terminus. However it is unclear as to why there should be peaks of BLAST hit start points at the termini. A process analogous to prokaryotic translation initiation would explain the peak for reverse orientation hits with the 18S rRNA but it does not however explain the peaks for other rRNAs or for hits in the forward orientation.

Altered search parameters

Tests were made to determine whether the results of the BLAST searches depended on the search parameters. The lower the Expect value, the less likely the BLAST hit is to be random. Short sequences, and sequences with a low sequence identity, would be reported with a high Expect value. The value of 10,000 was chosen, as it is the highest value that can be used. This was used so as many results as possible would be reported. A 10-fold reduction of this value to 1,000 leads to virtually no BLAST hits. As was expected because the smaller the database the larger the expect value will be for the same match.

A different way of sorting the data is to set the Expect value as high as possible to obtain as many BLAST hits as possible. Any results that are below a certain sequence identity are then discarded. A value of 75% sequence identity was chosen because known IRES in *S. cerevisiae* have a sequence identity to the 18S rRNA similar to this value (Zhou, Edelman et al. 2001). This however has a very similar effect to reducing the Expect value – most of the BLAST hits are discarded.

Discussion

One kilobase might be considered too long for a notional UTR because the average sequence length in the 5'-UTR dataset is 143 bp in size. However there is great variation in the lengths of the untranslated region. The standard deviation of the lengths is 234 bp (to 0 d.p.) and thus, assuming a Gaussain distribution 68% of the points lie within one standard deviation of the mean, 95% within two standard deviations and more than 99% within

three. Thus to be sure of accounting for most (at least 95%) of the 5'-UTRs a sequence of at least 143 + (2 x 234) bp = 611 bp should be used.

The use of a random 1 Kb sequence dataset and a dataset of intergenic regions as controls also help to determine if the results generated when using the 1 Kb upstream sequences are producing unusual results. The randomly generated sequences provide data on sequences that contain no biologically significant similarities to rRNA sequences. While the intergenic sequences provide the same data as the upstream sequences but without any influence from any bias caused by coding regions of the genome.

Three methods were used to analyse the results: the number and orientation of BLAST hits per sequence; the distribution of the number of BLAST hits; and the location on the rRNA of the BLAST hits. No evidence was obtained for IRES sites within the genome being a significant fraction of the UTRs.

The use of BLAST search as a predictive method for locating putative IRES yields no evidence to suggest that homology between the 18S rRNA and the mRNA is a major factor in internal translation initiation. However, it may well be important in a few specific cases and for which there is both experimental evidence (Paz, Abramovitz et al. 1999; Paz and Choder 2001; Zhou, Edelman et al. 2001) and an existing analogy with prokaryotic translation initiation (Mauro and Edelman 1997; Hu, Tranque et al. 1999). The genes with known 5'-UTR IRES activity have homology to 18S rRNA sequences, and functional IRES in *S. cerevisiae* have a high degree of homology to the 18S rRNA (Zhou, Edelman et al. 2001). The SIRES sequence (Paz, Abramovitz et al. 1999), which can support internal initiation in *S. cerevisiae*, was used in a BLAST search against the sequence databases set up for this investigation. However this resulted in almost no BLAST hits (data not shown). Aligning the SIRES sequence against the 18S rRNA sequence did however show that these two sequences are able to base pair with each other (diagram 5-5). Sequences were aligned using CLUSTAL W (1.82) on MAGI (Multiple Alignment General Interface) on the HGMP website (http://www.hgmp.mrc.ac.uk). The sequences were aligned using the

default settings of CLUSTAL W on this site. As the SIRES sequence has been shown to be an active IRES in *S. cerevisiae* this is another piece of evidence implicating homology with the 18S rRNA to IRES mediated translation initiation.

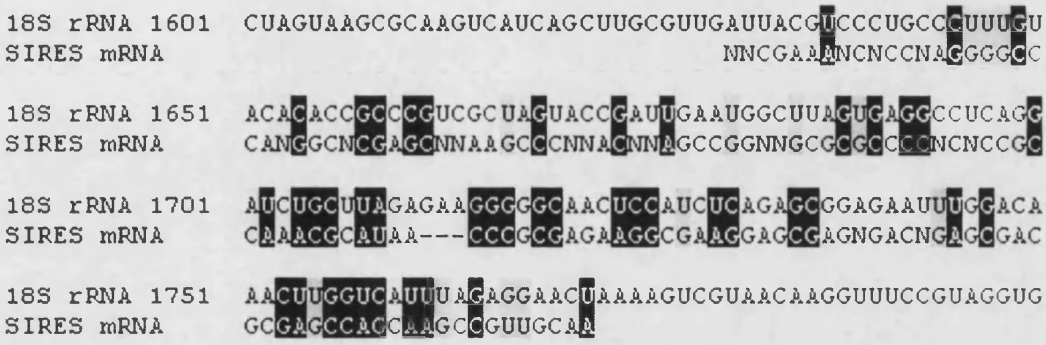


Diagram 5-5: Extensive base pairing can occur between the SIRES sequence and the 3' terminal end of the 18S rRNA. Base pairs highlighted in black indicate a Watson-Crick base pair, grey indicates a G:U "wobble" base pairing The rRNA sequence is arranged 5' to 3' (left to right), while the SIRES sequence is arranged 3' to 5'.

IRES sites have been characterised in both eukaryotic and viral systems. However in these cases the DNA sequence and lengths of these sites very divergent, and even function using different mechanisms (Kozak 1999; Martínez-Salas, Ramos et al. 2001; Pestova, Kolupaeva et al. 2001). The secondary structure of the RNAs has been suggested to play a very significant role. If it is the secondary structure of the 5'-UTR which plays a more pivotal role it would explain why no evidence of IRES could be found by analysing the sequence of 5'-UTRs. It will also make scanning large numbers of sequences for signs of potential IRES much more complicated.

WU-BLAST was used in the analysis because it allowed a large number of sequences in a specially constructed database to be searched for homology to a query

sequence. It could be set to allow specific base mismatches, which were used to account for G-U base pairing. Some of the BLAST hits returned were very long, with not a particularly high level of sequence identity (only 50% to 60%). For the IRES already discovered, sequence identity (including G-U base pairs) was 84% (Zhou, Edelman et al. 2001). Such matches may be of more use for predicting IRES if they were split into multiple shorter matches with a higher sequence identity.

The predictive ability of using BLAST to identify potential IRES may be dramatically improved if instead of using the *Expect* value of a match as a parameter, the percentage identity of a match is used instead. There are BLAST programs available which have this option, such as Mega BLAST (NCBI) but they are unable to allow specific base mismatches and therefore cannot take into account GU base pairing. We conclude that there is no evidence that IRES sequences perform a significant role in the genes that when deleted cause rapid death in stationary phase.

CHAPTER 6

DISCUSSION AND CONCLUSIONS

The screen for essential stationary phase genes

No large-scale, genome-wide screen can ever be 100% efficient in its ability to detect all potential candidate strains (Grunenfelder and Winzeler 2002), and this screen is no different. For example, genes that code for proteins that are part of redundant systems will not be detected because when one system is disrupted another system completes the task. To produce a visible phenotype a strain with more than one gene knockout must be constructed.

Another example is when genes cause a loss in viability in stationary phase when they are knocked out but this loss in viability might not be enough to pass the threshold used in the screening. Strains carrying a deletion in *HSP104* are known to suffer from a marginally reduced viability in stationary phase (Werner-Washburne, Braun et al. 1993). However, this gene was not identified in this screen for ORF knockouts because the threshold was set at a value of at least a 10^4-fold reduction in viability after 90 days, much more than is seen with the *HSP104* knockout.

Genes that are essential for vegetative growth will also not be identified. By their very nature, essential genes are difficult to test because deleting the gene is fatal. This limitation may now be avoidable. There are constructs under the control of *TET* promoters whose functions can be down regulated rather than fully repressed (Belli, Gari et al. 1998). These promoters can be used to allow full expression of the gene permitting normal growth and then the promoter can be turned down to allow expression of the gene for the test conditions. Another way of testing these genes would be to fuse the essential gene to a 'heat-inducible-degron' cassette (Kanemaki, Sanchez-Diaz et al. 2003). Using this method allows the cell to grow normally at permissive temperatures and then, when the cells are

exposed to the restrictive temperature, the target protein is degraded. This is unlike protein production control using a *TET* promoter, where the protein may exist in the cell for a long time after transcription of the gene has been halted. However, this does presume that in stationary phase protein degradation would occur at a similar rate.

Another example is the RAS/cAMP and TOR signalling pathways. If members of the TOR pathway, for example *PDE* genes or *CDC25*, are deleted the cell will constitutively enter stationary phase which means that the cell will fail to grow even when there are nutrients present, making it impossible to test for a stationary phase defect. These genes have already been identified as being part of the signalling pathways used when adapting to stationary phase but for genes not known to be involved with stationary phase, this may lead to them not being identified as being involved with stationary phase using methods described in this study.

In addition to these cases the lack of complete repetition during the screening of the entire SGDP gene knockout set, lack of some gene knockout strains in the set and contamination would almost certainly have produced some false negative results. There are also those cases during the first round of screening where the two haploid strains tested gave conflicting results. These ORF strains were not classified as not having a stationary phase defect, however they were not used in further testing potentially making some of these appear as false negatives. To counter this potential problem such strains should be re-screened to provide repetition of the results. However the identification of every gene associated with stationary phase is actually not required in a large-scale screen. If a large enough fraction of the genes are identified, those genes that were missed in the screen may well be identified during further research into those genes that were identified.

Another limitation with the method of screening used in this investigation is that it cannot experimentally determine the difference between primary and secondary effects of the ORF knockouts. The screen will identify those genes that code for proteins essential for

a process. It will also include genes whose proteins affect the function of those essential proteins, such as post-translational modification proteins.

Overall Conclusion

The single hypothesis that best unifies the results relates to energy generation and energy use. A cell in stationary phase has to face the same adverse conditions as a cell not in stationary phase. There is however a great difference in the energy and substrate stores that the cells can draw upon. When growing exponentially a cell can adapt to stresses and replace any damage. As there is no major restriction imposed by low energy levels (caused by the lack of a carbon source in the medium) anything that is required for adaptation/repair may be synthesised or imported from the medium. In stationary phase the reserves of energy and material that a cell can call upon to support these same functions is more limited. Furthermore during stationary phase transcription and translation rates are significantly reduced, which limits the production of any new proteins (in addition to the large energy and substrate requirements of protein synthesis). Once the reserves run out the cell may be unable to adapt to any stresses or repair any damage thus causing the cell to die.

Consequently anything that reduced the demand upon the energy/material that the cell managed to store as it entered stationary phase would make it easier for a cell to survive stationary phase. This could be achieved either directly or indirectly. Anything that directly affects the levels of stored energy and substrate in the cell will affect the cell's ability to survive stationary phase. Increasing the cellular reserves provides a larger resource for the cell to use in repair and adaptation, thus allowing it to survive in stationary phase for longer periods of time, while depleting these reserves causes the opposite effect. This can be seen in those strains for which the levels of glycogen and trehalose were tested. These substances are used to store energy in the cell but positive strains tested either

did not have these substances, or were at a level too small to be detected, and could not survive long-term stationary phase.

The other way to reduce the use of cellular stores is to lower the demand for them by active processes. This can be achieved by making the processes more efficient, allowing them to do more with the same amount of material or by reducing the need for any adaptation or repair by the cell. An example of this is any condition that provides extra protection against or reduces oxidative damage (such as the presence of reduced glutathione in the medium). Under these conditions the glutathione in the medium will reduce the oxidative damage the cell suffers. This allows energy/material that would have been used to repair oxidative damage to be used for other processes essential for stationary phase viability. The converse of this will reduce the ability of a cell to maintain viability in stationary phase, which is why gene knockouts that affect processes such as oxidative protection and protein synthesis cause cell death during long-term stationary phase

Growth media

Loss of viability in stationary phase was observed to be worse when strains were grown in SC medium compared to growth in YPD medium. This suggested that the less rich the medium during stationary phase the lower the strain's viability would be after long-term stationary phase. If this were the case incubation through stationary phase with water as the medium would result in an even larger loss of viability. Some strains have already been tested with water as the medium during stationary phase (loss of viability experiments, chapter 3). When incubated in stationary phase in water rather than exhausted medium these strains showed a rate of loss of viability that was the same if not faster than when incubated in YPD medium (Table 3-9). These results seem to confirm the prediction that strains incubated through long-term stationary phase in water, rather than spent medium, will lose viability at a faster rate. However this is a small sample of ORFs, and there are no data on medium of varying richness. If the predictions held true then all the

ORFs shown to cause a stationary phase defect should also act in this manner under these conditions.

During short-term stationary phase incubation, use of water as the medium rather than spent growth medium, reduces the loss of viability. The lack of this effect over the long term could be because in spent medium essential substances can still easily be imported into the cell from the medium. As in spent medium only one essential nutrient will have been completely exhausted. However incubation in water forces the cells to synthesise everything they require which will use up stores of energy and material at a faster rate and effectively starve the cell to death. This effect would not be immediately apparent, as the cell's internal stores would first be have to exhausted. This conclusion could be tested by growth in various media from YPD to a minimal medium. If long-term survival depended on the presence of substances in the medium the less rich the medium, the less viable the culture would be after long-term stationary phase. The use of various media might also allow the identification of essential extrinsic nutrients.

The initial results of the screens with YPD and YP(0.3D) media suggested that calorie restriction might be responsible for the rescue of the stationary phase defect when strains are grown in low glucose medium. However the fact that a lot of the ORF knockout strains are respiratory deficient, can be rescued by the addition of reduced glutathione, and that this effect is not observed when SC based medium is used suggests that this is not the case. While calorie restriction itself can be ruled out, a change associated with calorie restriction that is elicited by the conditions that cause calorie restriction to occur could be causing a change in the cell that protect the cell during stationary phase.

Respiration, oxidative stress, and the mitochondria

It will be very difficult, if not impossible, to completely separate these aspects because mitochondria are the sites of respiration in *S. cerevisiae* and respiration is the major cause of oxidative damage. Therefore anything that affects the mitochondria is very

likely to affect the ability of the cell to respire and in turn affect the degree of oxidative damage. The majority of the ORFs identified in this screen code for proteins that are respiratory or mitochondrial in function. This makes it obvious that the mitochondrion or something related to it (such as respiration) is highly important for stationary phase survival.

While the majority of the genes from the *S. cerevisiae* genome have been screened for a stationary phase essential phenotype there are still genes that have not been tested. The majority of these genes are genes for which there was no complete knockout available or no result could be obtained due to conflicting results or contamination. There are also those genes that are found on the mitochondrial genome. These genes were not part of the set of gene knockouts used in this study. But considering the apparent importance of the mitochondria and/or processes associated with it knockouts for these genes should be screened as well.

Respiration is an obvious process to consider, as this is probably the most important function of the mitochondria. It is also the cellular process that produces the most superoxide species (Longo, Gralla et al. 1996; Longo, Liou et al. 1999) and a relatively large number of proteins from the respiratory chain were identified as being essential for stationary phase survival. The ability of a cell to respire is not however essential for a cell to be able to survive long-term stationary phase, which was proved by the retesting of all available respiratory incompetent gene knockout strains. Even if respiration is not essential, it is a very important process as many gene knockout strains that cause the loss of respiration also prevent survival during long-term stationary phase.

SOD1, SOD2 and *GSH1* are induced during the shift from fermentative to oxidative metabolism, which makes the cells more resistant to H_2O_2 and to tert-butylhydroperoxide, and indicates that the antioxidant defences of *S. cerevisiae* are under carbon catabolite control. This is independent of *YAP1* and the *YAP1*-mediated stress response via *STRE* (Stress Response Element) (Maris, Kern et al. 2000; Maris, Assumpção et al. 2001). It has

been noted that longevity mutations postpone superoxide toxicity and mitochondrial damage, though increased antioxidant protection is only partially responsible for the major extension of lifespan caused by signal transduction mutations (Longo and Fabrizio 2002). A possible reason that not all respiratory deficient strains die during long-term stationary phase could be the generation of reactive oxygen species. The removal of different components of the respiratory chain could lead to different levels of reactive oxygen species generation thus leading to different levels of cellular damage and a different rate of loss of viability.

The glutaredoxin coding ORF YPL059w (*GRX5*) was identified. *GRX5* is a member of a glutaredoxin subfamily with *GRX3* and *GRX4*, which has significant sequence difference from the *GRX1/GRX2* glutaredoxin subfamily. Unlike other GRX proteins, Grx5p is part of mitochondrial machinery involved in synthesis/assembly of iron-sulphur centres. Thus knocking out this ORF would not only affect the antioxidant defences of the cell but also be detrimental to the proper function of respiration. Why this member of the glutaredoxins family should be more important than any of the other members is not obvious. It could be that *GRX5* is upregulated, while the other *GRX* genes are down regulated on adaptation to starvation making *GRX5* important for survival during stationary phase in a similar way to the situation with *UBI4*.

A properly functioning respiratory chain seems to be very important for stationary phase survival. In particular ubiquinone/co-enzyme Q and cytochrome c seem to be particularly important. Three ubiquinone biosynthetic genes (*COQ2*, *4* and *5*) were identified as being essential for stationary phase. While the exact function of *COQ4* is still unknown, the functions of *COQ2* and *COQ5* in ubiquinone biosynthesis have been identified (Diagram 6-1). Since these genes are essential for the maintenance of viability in stationary phase it suggests that ubiquinone is an essential factor. If this is the case the other proteins in the ubiquinone biosynthetic pathway (*COQ3*, *COQ6* and *CAT5*) (diagram 6-1) should also be essential. However, these genes were not positively identified as being

essential for stationary phase. *COQ3* and *COQ6* (YOL096c and YGR255c) both gave anomalous results, the MAT**a** strain retaining viability during stationary phase while the MATα strain lost viability. During the screen of the SGDP set of gene knockouts those cases where contradictory results were found, such as in the case of *COQ3 & 6* were not retested. The *CAT5* (YOR125c) gene knockout was not available in the SGDP set of ORF knockouts. The *CAT5* gene knockout should be created and all three gene knockouts should be retested. If ubiquinone is essential for stationary phase, over expression of these genes may result in increased viability in stationary phase, and if the levels of ubiquinone can be increased it may be able to rescue the single gene knockouts.

DMRL synthase (YOL143c/*RIB4*) may affect stationary phase in a similar way to the *COQ2*, *4* and *5* genes. DMRL is a precursor of riboflavin and FAD. $FADH_2$ feeds electrons into the respiratory chain at ubiquinone (diagram 6-2). The genes *RIB5*/YBR256c, *FMN1*/ YDR236c and *FAD1*/YDL045c convert DMRL to riboflavin, FMN and finally FAD. These genes were not identified in the screen for stationary phase essential genes as they were not part of the SGDP set of ORF knockouts. Knockouts of these genes should be created and tested to help confirm the importance of riboflavin/FAD. If knockouts of these genes were found to be essential it could be concluded that riboflavin and/or FAD are essential for stationary phase. This conclusion could be further confirmed if the cellular levels of riboflavin and or FAD could be increased (e.g. by incubating the cells in a medium containing riboflavin or FAD). Artificially increasing the levels of these chemicals should be able to rescue gene deletions of the DMRL to FAD pathway and it is possible that strains that over-express these genes might be able to survive stationary phase better than the wild type.

Hexaprenyl-diphophate + p-hydroxybenzoate

↓ *COQ2*

3-hexaprenyl-4-hydroxyenoate

↓

3,4-dihydroxy-5-hexaprenylbenzoate

↓ *COQ3*

3-mthoxy-4-hydroxy-5-hexaprenylbenzoate

↓

2-hexaprenyl-6-methoxyphenol

↓ *COQ6*

2-hexaprenyl-6-methoxy-1,4-benzoquinol

↓ *COQ5*

2-hexaprenyl-3-methyl-6-methoxy-1,4-benzoquinol

↓ *CAT5*

2-hexaprenyl-3-methyl-5-hydroxy-6-methoxy-1,4-benzoquinol

↓ *COQ3*

ubiquinol (30)

↓ *spontaneous*

ubiquinone (30)

Diagram 6-1: Ubiquinone biosynthesis pathway, based on a diagram from the Saccharomyces Genome Database (SGD 2002).

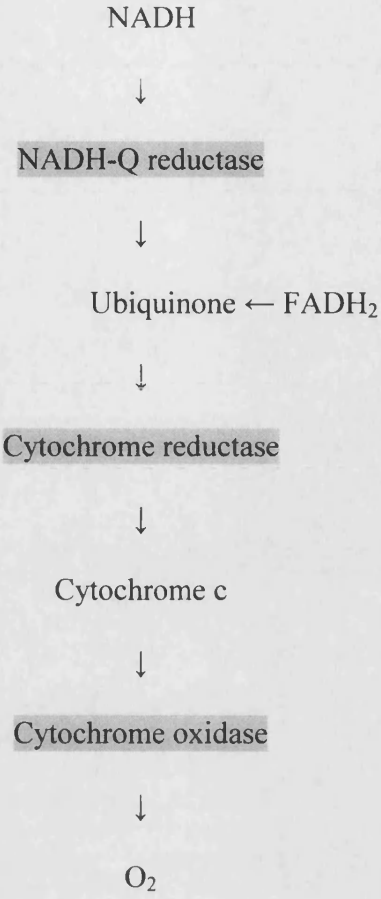


Diagram 6-2: Sequence of electron carriers in the respiratory chain. The proton pump protein complexes are highlighted in grey (based on figure 21.5, (Stryer 1995).

All the cytochrome c oxidase associated proteins identified by this investigation (*COX9, 10, 14, 18* and *19*) are all involved in cytochrome c oxidase biogenesis. In addition there were several genes identified that are involved in the transcription and translation of *COB* and *COX* genes. In addition to this the genes *COX1*, *COX2*, *COX3* and *COB1* are all coded for on the mitochondrial genome (table 6-1). While knockouts of these genes were not directly tested, knockouts of mitochondrial ribosomal proteins were tested. Without properly functioning mitochondrial ribosomes these proteins coded for by these genes cannot be made, and are effectively knocked out as well. Thus the loss of viability seen with the mitochondrial ribosome gene knockouts could not be directly due to the loss of ribosome function, rather it was an effect on cytochrome c oxidase function. Cytochrome c

oxidase (respiratory complex IV) is a polypeptide consisting of 12 proteins. Only one of the subunits was directly identified even though several of the subunits stop respiration when knocked out. This suggests that cytochrome c oxidase may have some function other than being part of the electron transport chain in respiration, which could be an antioxidant role (Korshunov, Krasnikov et al. 1999).

Non-respiratory consumption of oxygen is mainly localised in the mitochondria, and is typically used when the respiratory chain is impaired or absent. Sterol biosynthesis accounts for roughly ¼ of this, with the rest being unknown pathway(s). These activities are associated with a massive production of $O_2^{\cdot -}$ and to lesser extent H_2O_2. (Rosenfeld, Beauvoit et al. 2002). Thus the disruption of the respiratory chain could lead to an increase of superoxide production, which causes an increased loss of viability in stationary phase. Any strain without mitochondrial function (e.g. rho^0 strains, or after antimycin A treatment) are more sensitive to H_2O_2 (Grant, MacIver et al. 1997; Aguilaniu, Gustafsson et al. 2001) suggesting an impairment of the antioxidant defences or an increase in oxidative damage to a level greater than that with which the defences can cope. However, this is probably not the case since if it were then any mutation of proteins that make up the respiratory electron transport chain that abolished the ability to respire would cause viability to be lost in stationary phase. Many of the genes involved in the respiratory electron transport chain can be knocked out without a loss of stationary phase viability. Furthermore respiratory deficient strains are able to respond properly to oxidative stress (Grant, MacIver et al. 1997) and there is no correlation of respiratory deficiency and rescue by reduced glutathione.

ORFs of unknown function

Several of the ORFs identified have not yet had a function assigned to them. There are many software programs available for the prediction of similarity, function, localisation and other properties of a protein. By looking at the results of many different analyses it is

sometimes possible to make educated guesses about the role and nature of the protein in question. The program PIX (http://www.hgmp.mrc.ac.uk) submits the protein sequence to many different analyses using reasonable default values and returns the results of all the analyses together. Using PIX on those genes that are essential for stationary phase but have no assigned function suggested that gene YDR065w codes for a nucleotide binding protein. BLAST results suggest it has a region of similarity to zebrafish (*Danio rerio*) gridded kidney cDNA clone 4759507-5 and to human heptad centromeric protein E. It also has one of the five Prosite motifs for both the ribosomal HS6 and prokaryotic DNA topoisomerase I signatures. The results for YGR102c, YGR150c, YMR098c, YOR305w, YPR099c (which is probably not a gene (Wood, Rutherford et al. 2001)), and YPR116w are not clear enough for any conclusion about a possible function to be made.

The failure to assign functions to them makes it possible that they have a specific stationary phase-related function. This however makes it no easier to conclude what part in stationary phase the proteins play. ORF YDR065w is an exception. The similarity that it has with other proteins suggests that its function involves binding to DNA. This suggests that it could function as a transcription factor under the control of one or more signalling pathways to directly influence the transcription of genes in a manner required for stationary phase. If more evidence is found that the protein coded for by this gene is a transcription factor a band-shift assay using possible the possible regulatory sequences identified in during the bioinformatic investigations during this study and the purified protein could be undertaken. This could potentially confirm that the protein coded for by YDR065w is a transcription factor and that one or more of the sequences identified are real transcription factor binding sites.

ORFs of known function

The proteins coded for by the ORFs YDR507c (*GIN4*), YDR523c (*SPS1*), and YPL031c (*PHO85*) are all possible signal pathway components. Gin4p interacts with cell

division control proteins Cdc3p, Cdc10p, Cdc11p, Cdc12p (all of which are essential for cytokinesis), Nap1p (nucleosome assembly protein I) and septin Sep7p. Pho85p interacts with Pcl2p, 5p 6p, 7p, 8p, 9p and 10p) and and Clg1p (components of the cyclin-dependent protein kinase holoenzyme complex), YDL246c, and YNL201c. The *CDC*, *PCL*, and *CLG1* genes are used to control the cell cycle. Therefore Gin4p and Pho85p could be involved with linking nutrient sensing signal pathways to the cell cycle control system. Though the interaction data do not indicate whether these two genes control or are controlled by the cell cycle machinery. YNL225c (*CNM67*) could also be involved in regulating the cell cycle. Like the proteins encoded by *CDC3, 10, 11* and *12* it is part of the cytoskeleton and is involved in cytokinesis. It is part of the spindle pole body, unlike the *CDC* genes that form part of the septin ring. When examined by microscopy these ORF knockout strains (in stationary phase) did not appear to be different from any other ORF knockout strain or *S. cerevisiae* FY1679. Therefore the loss of these genes does not appear to be affecting the point during the cell cycle where stationary phase cells arrest growth. If they were affecting this process cells in cultures of these three gene knockouts would be entering stationary phase at a random point during the cell cycle and the cultures would appear to have a larger proportion of cells that are budding.

The protein encoded by YLR260w (*LCB5*) is a sphingolipid long chain base (LCB) kinase and is used for sphingolipid biosynthesis. It has been shown that mutants of the *OPI3* gene have impaired stationary phase viability (McGraw and Henry 1989). Opi3p is involved in phophatidylcholine biosynthesis, and it has been suggested that membrane structure is important in maintaining viability in stationary phase (McGraw and Henry 1989; Werner-Washburne, Braun et al. 1993). As well as being a structural component of the cell membrane, sphingolipids have been implicated in regulating cell growth, differentiation, angiogenesis, apoptosis, and senescence (Obeid, Okamoto et al. 2002). Thus mutants of *LCB5* could affect stationary phase viability by affecting the structure of the cell membrane, or by interfering with a signalling pathway.

YPL045w (*VPS16*) is a vacuolar sorting protein. The defect caused by this gene knockout might be because in the knockout the cell cannot recycle proteins properly. It is known that most of the ubiquitin-proteasome pathway is down regulated during stationary phase (Bajorek, Finley et al. 2003). Therefore proteins that are being sent to be recycled, but are wrongly targeted, will not be recycled. This would cause a cell to receive less energy/material from the recycling of old protein material. This would be detrimental to a cell under strict energy and material limiting conditions as the limited cellular stores of energy and material will be consumed at a faster rate causing the cell to starve to death quicker. There is no obvious reason as to why this component is more important than any other vacuolar sorting component. Of all the *UBI* genes only *UBI4* is induced during stationary phase (Finley, Özkaynak et al. 1987; Fraser, Luu et al. 1991; Werner-Washburne, Braun et al. 1993). If Vps16p is found to interact with ubiquitin (a protein degradation signal), the altered ubiquitination of proteins could explain the importance of Vps16p during stationary phase.

YNL139c (*RLR1*) and YDR138w (*HPR1*) play a role in transcription elongation, as part of THO complex that is crucial for transcription of certain coding regions. Exactly how this affects stationary phase is not obvious. The other two subunits of the THO complex (*THP2*/YHR167w and *MFT1*/YML062c) were found not to be essential for stationary phase. These two genes should be retested, as the stationary phase non-essential phenotype may be a false negative result.

YJR122w (*CAF17*) is an associated factor of the *CCR4* transcriptional complex. The *CCR4* complex is a transcriptional regulatory complex and affects transcription both positively and negatively. Both these complexes could be activated or deactivated by one or more of the signal pathways involved in nutrient sensing thereby affecting transcription and setting the cell up for stationary phase. If the genes regulated by, and the signal pathways upstream of, the THO and CCR4 complexes can be identified and linked to stationary phase a better understanding of the control of stationary phase could be gained.

A large number of identified genes encoded ribosomal proteins. In particular the large subunit of the mitochondrial ribosome (Table 2-3) was over-represented in the set. This suggests that translation in the mitochondria is important for the maintenance of viability in stationary phase. Several proteins from the small subunit of the mitochondrial ribosome were also identified but only one protein from both subunits of the cellular ribosome was identified which indicates that it is some aspect of translation in the mitochondria and not translation in general that is important. There are 28 genes on the mitochondrial genome, eleven of which have an unknown molecular function (table 6-1). Of the rest, approximately half have a nuclease activity or bind RNA and one is part of the ribosome. The rest have either cytochrome c oxidase activity or are part of a hydrogen ion pump. Several times during this investigation the process of respiration has been found to be important for the maintenance of viability during stationary phase and this process would be severely disrupted by a loss of mitochondrial translation. Therefore it is probable that deletion of the mitochondrial ribosomal genes is not primarily responsible for the loss in viability in stationary phase but causes the loss of viability by preventing the production of proteins that are essential for stationary phase survival

Systematic Name (gene name)	Function
Q0045 (COX1), Q0250 (COX2), Q0275 (COX3)	cytochrome-c oxidase activity
Q0105 (COB1)	ubiquinol-cytochrome-c reductase activity
Q0050 (AI1), Q0060 (AI3), Q0065 (AI4), Q0070 (AI5_α), Q0160 (SCEI)	endonuclease activity
Q0130 (ATP9)	hydrogen ion transporter activity
Q0080 (ATP8), Q0085 (ATP6)	hydrogen-transporting ATP synthase activity, rotational mechanism
Q0110 (BI2)	nuclease activity
Q0055 (AI2), Q0115 (BI3), Q0120 (BI4)	RNA binding
Q0140 (VAR1)	structural constituent of ribosome
Q0010 (ORF6), Q0017 (ORF7), Q0032 (ORF8), Q0075 (AI5_β), Q0092 (ORF5), Q0142 (ORF9), Q0143 (ORF10), Q0144 ,Q0182 (ORF11), Q0255 (RF1), Q0297 (ORF12)	molecular_function unknown

Table 6-1: Genes on the *S. cerevisiae* mitochondrial genome. Data in table is available on the SGD website (Cherry, Ball et al. 2002).

Bibliography

Achstetter, T., C. Ehmann, et al. (1983). Proteolysis in Eukaryotic Cells: Aminopeptidases and Dipeptidyl Aminopeptidases of Yeast Revisited. *Archives of Biochemistry and Biophysics* **226**(1): 292-305.

Aguilaniu, H., L. Gustafsson, et al. (2001). Protein oxidation in G_0 cells of *Saccharomyces cerevisiae* depends on the state rather than rate of respiration and is enhanced in pos9 but not yap1 mutants. *Journal of Biological Chemistry* **276**(38): 35396-35404.

Ashrafi, K., D. Sinclair, et al. (1999). Passage through stationary phase advances replicative aging in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America* **96**(16): 9100-9105.

Bajorek, M., D. Finley, et al. (2003). Proteasome disassembly and downregulation is correlated with viability during stationary phase. *Current Biology* **12**(13): 1140-1144.

Baxevanis, A. D. (2000). The Molecular Biology Database Collection: an online compilation of relevant database resources. *Nucleic Acids Research* **28**(1): 1-7.

Belli, G., E. Gari, et al. (1998). Functional analysis of yeast essential genes using a promoter- substitution cassette and the tetracycline-regulatable dual expression system. *Yeast* **14**(12): 1127-1138.

Bertram, P. G., J. H. Choi, et al. (2000). Tripartite regulation of Gln3p by TOR, Ure2p, and phosphatases. *Journal of Biological Chemistry* **275**(46): 35727-35733.

Bertram, P. G., J. H. Choi, et al. (2002). Convergence of TOR-nitrogen and Snf1-glucose signaling pathways onto Gln3. *Molecular and Cellular Biology* **22**(4): 1246-1252.

Broach, J. R. (1991). Ras Genes in *Saccharomyces cerevisiae* - Signal Transduction in Search of a Pathway. *Trends in Genetics* **7**(1): 28-33.

Brown, A. J. P. and M. F. Tuite (1998). Appendix II: Yeast Growth. Yeast Gene Analysis. A. J. P. Brown and M. F. Tuite. *London, Academic Press.* **26:** 471-477.

Bugeja, V. C., J. R. Piggott, et al. (1982). Differentiation of *Saccharomyces cerevisiae* at Slow Growth- Rates in Glucose-Limited Chemostat Culture. *Journal of General Microbiology* **128**(NOV): 2707-2714.

Cannon, J. F. and K. Tatchell (1987). Characterization of *Saccharomyces cerevisiae* Genes Encoding Subunits of Cyclic Amp-Dependent Protein-Kinase. *Molecular and Cellular Biology* **7**(8): 2653-2663.

Carlson, M. (1999). Glucose repression in yeast. *Current Opinion in Microbiology* **2**(2): 202-207.

Cherry, J. M., C. Ball, et al. (2002). *Saccharomyces* Genome Database. **2002.**

Chu, S., J. DeRisi, et al. (1998). The transcriptional program of sporulation in budding yeast. *Science* **282**(5389): 699-705.

Cyrne, L., L. Martins, et al. (2003). Regulation of antioxidant enzymes gene expression in the yeast *Saccharomyces cerevisiae* during stationary phase. *Free Radical Biology and Medicine* **34**(3): 385-393.

de la Cruz, B. J., S. Prieto, et al. (2002). The role of the 5 ' untranslated region (UTR) in glucose-dependent mRNA decay. *Yeast* **19**(10): 887-902.

de Nobel, J. G., F. M. Klis, et al. (1990). The Glucanase-Soluble Mannoproteins Limit Cell-Wall Porosity in *Saccharomyces cerevisiae. Yeast* **6**(6): 491-499.

DeRisi, J. L., V. R. Iyer, et al. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**(5338): 680-686.

Dorokhov, Y. L., M. V. Skulachev, et al. (2002). Polypurine (A)-rich sequences promote cross-kingdom conservation of internal ribosome entry. *Proceedings of the National Academy of Sciences of the United States of America* **99**(8): 5301-5306.

Dorsman, J. C., M. M. Doorenbosch, et al. (1989). An Ars Silencer Binding-Factor Also Activates 2 Ribosomal- Protein Genes in Yeast. *Nucleic Acids Research* **17**(13): 4917-4923.

Drebot, M. A., C. A. Barnes, et al. (1990). Genetic Assessment of Stationary Phase for Cells of the Yeast *Saccharomyces cerevisiae*. *Journal of Bacteriology* **172**(7): 3584-3589.

Drebot, M. A., G. C. Johnston, et al. (1987). A Yeast Mutant Conditionally Defective Only for Reentry into the Mitotic Cell-Cycle from Stationary Phase. *Proceedings of the National Academy of Sciences of the United States of America* **84**(22): 7948-7952.

Eisen, M. B., P. T. Spellman, et al. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America* **95**(25): 14863-14868.

Elert, G. Diameter of a Yeast. **2003**.

Esposito, R. E. and S. Klapholz (1981). Meiosis and ascospore development. The molecular biology of the Saccharomyces life cycle and inheritance. J. N. Strathern, E. W. Jones and J. R. Broach. *Cold Spring Harbor, Cold Spring Harbor Laboratory Press*: 211-.

Ferea, T. L., D. Botstein, et al. (1999). Systematic changes in gene expression patterns following adaptive evolution in yeast. *Proceedings of the National Academy of Sciences of the United States of America* **96**(17): 9721-9726.

Finley, D., E. Özkaynak, et al. (1987). The Yeast Polyubiquitin Gene Is Essential for Resistance to High-Temperatures, Starvation, and Other Stresses. *Cell* **48**(6): 1035-1046.

Fraser, J., H. A. Luu, et al. (1991). Ubiquitin Gene-Expression: Response to Environmental-Changes. *Current Genetics* **20**(1-2): 17-23.

Fröhlich, K.-U. and F. Madeo (2001). Apoptosis in yeast: a new model for aging research. *Experimental Gerontology* **37**(1): 27-31.

Fuge, E. K., E. L. Braun, et al. (1994). Protein-Synthesis in Long-Term Stationary-Phase Cultures of *Saccharomyces cerevisiae*. *Journal of Bacteriology* **176**(18): 5802-5813.

Gachon, F., G. Gaudray, et al. (2001). The cAMP response element binding protein-2 (CREB-2) can interact with the C/EBP-homologous protein (CHOP). *Febs Letters* **502**(1-2): 57-62.

Gadal, O., D. Strauß, et al. (2001). A nuclear AAA-type ATPase (Rix7p) is required for biogenesis and nuclear export of 60S ribosomal subunits. *Embo Journal* **20**(14): 3695-3704.

Gailus-Durner, V., J. X. Xie, et al. (1996). Participation of the yeast activator Abf1 in meiosis-specific expression of the HOP1 gene. *Molecular and Cellular Biology* **16**(6): 2777-2786.

Gancedo, J. M. (1998). Yeast carbon catabolite repression. *Microbiology and Molecular Biology Reviews* **62**(2): 334-+.

Gasch, A. P., M. Huang, et al. (2001). Genomic expression responses to DNA-damaging agents and the regulatory role of the yeast ATR homolog Mec1p. *Molecular Biology of the Cell* **12**(10): 2987-3003.

Gasch, A. P., P. T. Spellman, et al. (2000). Genomic expression programs in the response of yeast cells to environmental changes. *Molecular Biology of the Cell* **11**(12): 4241-4257.

Gerrard, J. A., A. D. Sparrow, et al. (2001). Metabolic databases - what next? *Trends in Biochemical Sciences* **26**(2): 137-140.

Goffeau, A., B. G. Barrell, et al. (1996). Life with 6000 genes. *Science* **274**(5287): 546-&.

Görner, W., E. Durchschlag, et al. (2002). Acute glucose starvation activates the nuclear localization signal of a stress-specific yeast transcription factor. *Embo Journal* **21**(1-2): 135-144.

Granot, D. and M. Snyder (1991). Glucose Induces cAMP-Independent Growth-Related Changes in Stationary-Phase Cells of *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America* **88**(13): 5724-5728.

Granot, D. and M. Snyder (1993). Carbon Source Induces Growth of Stationary Phase Yeast-Cells, Independent of Carbon Source Metabolism. *Yeast* **9**(5): 465-479.

Grant, C. M., F. H. MacIver, et al. (1996). Glutathione is an essential metabolite required for resistance to oxidative stress in the yeast *Saccharomyces cerevisiae*. *Current Genetics* **29**(6): 511-515.

Grant, C. M., F. H. MacIver, et al. (1997). Mitochondrial function is required for resistance to oxidative stress in the yeast *Saccharomyces cerevisiae*. *Febs Letters* **410**(2-3): 219-222.

Grunenfelder, B. and E. A. Winzeler (2002). Treasures and traps in genome-wide data sets: Case examples from yeast. *Nature Reviews Genetics* **3**(9): 653-661.

Hampsey, M. (1997). A review of phenotypes in *Saccharomyces cerevisiae*. *Yeast* **13**(12): 1099-1133.

Hardwick, J. S., F. G. Kuruvilla, et al. (1999). Rapamycin-modulated transcription defines the subset of nutrient-sensitive signaling pathways directly controlled by the Tor proteins. *Proceedings of the National Academy of Sciences of the United States of America* **96**(26): 14866-14870.

Harris, N., M. MacLean, et al. (2001). Increasing *Saccharomyces cerevisiae* stress resistance, through the overactivation of the heat shock response resulting from defects in the Hsp90 chaperone, does not extend replicative life span but can be associated with slower chronological ageing of nondividing cells. *Molecular Genetics and Genomics* **265**(2): 258-263.

Henstock, M. R. and A. E. Wheals (2001). Internet sites useful for research and teaching with yeasts. *FEMS Yeast Research* **1**(1): V-VII.

Henstock, M. R. and A. E. Wheals (2002). Some useful internet sites for research and teaching with yeasts. *FEMS Yeast Research* **2**(1): I-IV.

Herman, P. K. (2002). Stationary phase in yeast. *Current Opinion in Microbiology* **5**(6): 602-607.

Hu, M. C.-Y., P. Tranque, et al. (1999). rRNA-complementarity in the 5 ' untranslated region of mRNA specifying the Gtx homeodomain protein: Evidence that base-pairing to 18S rRNA affects translational efficiency. *Proceedings of the National Academy of Sciences of the United States of America* **96**(4): 1339-1344.

Hurst, L. D., E. J. B. Williams, et al. (2002). Natural selection promotes the conservation of linkage of co- expressed genes. *Trends in Genetics* **18**(12): 604-606.

Ihmels, J., G. Friedlander, et al. (2002). Revealing modular organization in the yeast transcriptional network. *Nature Genetics* **31**(4): 370-377.

Ireland, L. S., G. C. Johnston, et al. (1994). A Member of a Novel Family of Yeast 'Zn-Finger' Proteins Mediates the Transition from Stationary-Phase to Cell-Proliferation. *Embo Journal* **13**(16): 3812-3821.

Jakubowski, W., T. Bilinski, et al. (2000). Oxidative stress during aging of stationary cultures of the yeast *Saccharomyces cerevisiae*. *Free Radical Biology and Medicine* **28**(5): 659-664.

Johnston, C. G., J. R. Pringle, et al. (1977). coordination of growth with cell division in the yeast *Saccharomyces cerevisiae*. *Experimental Cell Research* **105**: 79-98.

Jona, G., M. Choder, et al. (2000). Glucose starvation induces a drastic reduction in the rates of both transcription and degradation of mRNA in yeast. *Biochimica Et Biophysica Acta-Gene Structure and Expression* **1491**(1-3): 37-48.

Jones, E. W. (1984). The Synthesis and Function of Proteases in *Saccharomyces* - Genetic Approaches. *Annual Review of Genetics* **18**: 233-270.

Jørgensen, H., L. Olsson, et al. (2002). Fed-batch cultivation of baker's yeast followed by nitrogen or carbon starvation: effects on fermentative capacity and content of

trehalose and glycogen. *Applied Microbiology and Biotechnology* **59**(2-3): 310-317.

Kaida, D., H. Yashiroda, et al. (2002). Yeast Whi2 and Psr1-phosphatase form a complex and regulate STRE-mediated gene expression. *Genes to Cells* **7**(6): 543-552.

Kanemaki, M., A. Sanchez-Diaz, et al. (2003). Functional proteomic identification of DNA replication proteins by induced proteolysis in vivo. *Nature* **423**(6941): 720-724.

Kataoka, T., D. Broek, et al. (1985). DNA-Sequence and Characterization of the S-Cerevisiae Gene Encoding Adenylate-Cyclase. *Cell* **43**(2): 493-505.

Kolter, R., D. A. Siegele, et al. (1993). The Stationary-Phase of the Bacterial Life-Cycle. *Annual Review of Microbiology* **47**: 855-874.

Korshunov, S. S., B. F. Krasnikov, et al. (1999). The antioxidant functions of cytochrome c. *Febs Letters* **462**(1-2): 192-198.

Kozak, M. (1999). Initiation of translation in prokaryotes and eukaryotes. *Gene* **234**(2): 187-208.

Krause, S. A. and J. V. Gray (2002). The protein kinase C pathway is required for viability in quiescence in *Saccharomyces cerevisiae*. *Current Biology* **12**(7): 588-593.

Lagunas, R. (1986). Misconceptions About the Energy Metabolism of *Saccharomyces Cerevisiae*. *Yeast* **2**(4): 221-228.

Lalucque, H. and P. Silar (2000). In vivo labelling of functional ribosomes reveals spatial regulation during starvation in *Podospora anserina*. *BMC Genetics* **1**(1): 3-5.

Lillie, S. H. and J. R. Pringle (1980). Reserve carbohydrate metabolism in *Saccharomyces cerevisiae*: Responses to nutrient limitation. *Journal of Bacteriology* **143**(3): 1384-1394.

Lin, S.-J., M. Kaeberlein, et al. (2002). Calorie restriction extends *Saccharomyces cerevisiae* lifespan by increasing respiration. *Nature* **418**(6895): 344-348.

Longo, V. D. and P. Fabrizio (2002). Visions & reflections - Regulation of longevity and stress resistance: a molecular strategy conserved from yeast to humans? *Cellular and Molecular Life Sciences* **59**(6): 903-908.

Longo, V. D., E. B. Gralla, et al. (1996). Superoxide dismutase activity is essential for stationary phase survival in *Saccharomyces cerevisiae* - Mitochondrial production of toxic oxygen species in vivo. *Journal of Biological Chemistry* **271**(21): 12275-12280.

Longo, V. D., L.-L. Liou, et al. (1999). Mitochondrial superoxide decreases yeast survival in stationary phase. *Archives of Biochemistry and Biophysics* **365**(1): 131-142.

MacLean, M., N. Harris, et al. (2001). Chronological lifespan of stationary phase yeast cells; a model for investigating the factors that might influence the ageing of postmitotic tissues in higher organisms. *Yeast* **18**(6): 499-509.

Mahler, H. R. and P. S. Perlman (1972). Effects of mutagenic treatment by ethidium bromide on cellular and mitochondrial phenotype. *Archives of Biochemistry and Biophysics* **148**: 115-129.

Maris, A. F., A. L. K. Assumpção, et al. (2001). Diauxic shift-induced stress resistance against hydroperoxides in *Saccharomyces cerevisiae* is not an adaptive stress response and does not depend on functional mitochondria. *Current Genetics* **39**(3): 137-149.

Maris, A. F., A. L. Kern, et al. (2000). Glutathione, but not transcription factor Yap1, is required for carbon source-dependent resistance to oxidative stress in *Saccharomyces cerevisiae*. *Current Genetics* **37**(3): 175-182.

Martinez, A. Saccharomyces - Soil Microbiology: BIOL/CSES 4684. **2003**.

Martínez-Pastor, M. T. and F. Estruch (1996). Sudden depletion of carbon source blocks translation, but not transcription, in the yeast *Saccharomyces cerevisiae*. *Febs Letters* **390**(3): 319-322.

Martínez-Salas, E., R. Ramos, et al. (2001). Functional interactions in internal translation initiation directed by viral and cellular IRES elements. *Journal of General Virology* **82**: 973-984.

Mauro, V. P. and G. M. Edelman (1997). rRNA-like sequences occur in diverse primary transcripts: Implications for the control of gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **94**(2): 422-427.

McCarthy, J. E. G. (1998). Posttranscriptional control of gene expression in yeast. *Microbiology and Molecular Biology Reviews* **62**(4): 1492-1553.

McGraw, P. and S. A. Henry (1989). Mutations in the *Saccharomyces cerevisiae* Opi3 Gene - Effects on Phospholipid Methylation, Growth and Cross-Pathway Regulation of Inositol Synthesis. *Genetics* **122**(2): 317-330.

Mercier, G., Y. Denis, et al. (2001). Transcriptional induction of repair genes during slowing of replication in irradiated *Saccharomyces cerevisiae*. *Mutation Research-DNA Repair* **487**(3-4): 157-172.

Nestlbacher, R., P. Laun, et al. (2000). The influence of oxygen toxicity on yeast mother cell-specific aging. *Experimental Gerontology* **35**(1): 63-70.

Obeid, L. A., Y. Okamoto, et al. (2002). Yeast sphingolipids: metabolism and biology. *Biochimica Et Biophysica Acta-Molecular and Cell Biology of Lipids* **1585**(2-3): 163-171.

Pál, C. and L. D. Hurst (2003). Evidence for co-evolution of gene order and recombination rate. *Nature Genetics* **33**(3): 392-395.

Parrou, J. L. and J. François (1997). A simplified procedure for a rapid and reliable assay of both glycogen and trehalose in whole yeast cells. *Analytical Biochemistry* **248**(1): 186-188.

Paz, I., L. Abramovitz, et al. (1999). Starved *Saccharomyces cerevisiae* cells have the capacity to support internal initiation of translation. *Journal of Biological Chemistry* **274**(31): 21741-21745.

Paz, I. and M. Choder (2001). Eukaryotic translation initiation factor 4E-dependent translation is not essential for survival of starved yeast cells. *Journal of Bacteriology* **183**(15): 4477-4483.

Pérez-Torrado, R., J. V. Gimeno-Alcaniz, et al. (2002). Wine yeast strains engineered for glycogen overproduction display enhanced viability under glucose deprivation conditions. *Applied and Environmental Microbiology* **68**(7): 3339-3344.

Pestova, T. V., V. G. Kolupaeva, et al. (2001). Molecular mechanisms of translation initiation in eukaryotes. *Proceedings of the National Academy of Sciences of the United States of America* **98**(13): 7029-7036.

Piñon, R. (1978). Folded chromosomes in non-cycling yeast cells. Evidience for a characteristic G_0 form. *Chromosoma* **67**: 263-274.

Planta, R. J., P. M. Goncalves, et al. (1995). Global regulators of ribosome biosynthesis in yeast. *Biochemistry and Cell Biology-Biochimie Et Biologie Cellulaire* **73**(11-12): 825-834.

Plourde-Owobi, L., S. Durner, et al. (2000). Trehalose reserve in Saccharomyces cerevisiae: phenomenon of transport, accumulation and role in cell viability. *International Journal of Food Microbiology* **55**(1-3): 33-40.

Pringle, J. R. and L. H. Hartwell (1981). The *Saccharomyces cerevisiae* cell cycle. Molecular biology of the yeast Saccharomyces: Life cycle and inheritance. J. Broach, J. Strathern and E. Jones. *New York, Cold Spring Harbor Laboratory Press.*

Rosenfeld, E., B. Beauvoit, et al. (2002). Non-respiratory oxygen consumption pathways in anaerobically- grown *Saccharomyces cerevisiae*: evidence and partial characterization. *Yeast* **19**(15): 1299-1321.

Ross-Macdonald, P., P. S. R. Coelho, et al. (1999). Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature* **402**(6760): 413-418.

Roth, S. and H.-J. Schüller (2001). Cat8 and Sip4 mediate regulated transcriptional activation of the yeast malate dehydrogenase gene MDH2 by three carbon source-responsive promoter elements. *Yeast* **18**(2): 151-162.

Roy, A. and A. K. Ghosh (1998). Correlation between stationary phase survival and acid trehalase activity in yeast. *Biochimica Et Biophysica Acta-Molecular Cell Research* **1401**(3): 235-238.

Ruis, H. and C. Schüller (1995). Stress Signaling in Yeast. *Bioessays* **17**(11): 959-965.

Rupes, I. (2002). Checking cell size in yeast. *Trends in Genetics* **18**(9): 479-485.

Sachs, A. B. and G. Varani (2000). Eukaryotic translation initiation: there are (at least) two sides to every story. *Nature Structural Biology* **7**(5): 356-361.

Samokhvalov, V., V. Ignatov, et al. (2004). Reserve carbohydrates maintain the viability of *Saccharomyces cerevisiae* cells during chronological aging. *Mechanisms of Ageing and Development* **125**(3): 229-235.

Sandmeier, J. J., S. French, et al. (2002). RPD3 is required for the inactivation of yeast ribosomal DNA genes in stationary phase. *EMBO J.* **21**(18): 4959-4968.

Schmelzle, T. and M. N. Hall (2000). TOR, a central controller of cell growth. *Cell* **103**(2): 253-262.

SCPD (1998). The Promoter Database of Saccharomyces cerevisiae. http://cgsigma.cshl.org/jian/

SGD (2002). S. cerevisiae Pathway: ubiquinone biosynthesis. http://pathway.yeastgenome.org:8555/YEAST/new-image?type=PATHWAY&object=PWY3O-862&detail-level=2

Silljé, H. H. W., J. W. G. Paalman, et al. (1999). Function of trehalose and glycogen in cell cycle progression and cell viability in *Saccharomyces cerevisiae*. *Journal of Bacteriology* **181**(2): 396-400.

Slaughter, J. C. and T. Nomura (1992). Intracellular Glycogen and Trehalose Contents as Predictors of Yeast Viability. *Enzyme and Microbial Technology* **14**(1): 64-67.

Sobering, A. K., U. S. Jung, et al. (2002). Yeast Rpi1 is a putative transcriptional regulator that contributes to preparation for stationary phase. *Eukaryotic Cell* **1**(1): 56-65.

Spellman, P. T., G. Sherlock, et al. (1998). Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Molecular Biology of the Cell* **9**(12): 3273-3297.

Stevens, B. (1981). Mitochondrial structure. Molecular biology of the yeast *Saccharomyces*: Life cycle and inheritance. J. Broach, J. Strathern and E. Jones. *New York, Cold Spring Harbor Laboratory Press*: 471-504.

Stryer, L. (1995). Biochemistry. *New York, W. H. Freeman and Company*, ISBN: 0716720094.

Sudbery, P. E., A. R. Goodey, et al. (1980). Genes which control cell proliferation in the yeast *Saccharomyces cerevisiae*. *Nature* **288**: 401-404.

Tamai, K. T., X. D. Liu, et al. (1994). Heat-Shock Transcription Factor Activates Yeast Metallothionein Gene-Expression in Response to Heat and Glucose Starvation Via Distinct Signaling Pathways. *Molecular and Cellular Biology* **14**(12): 8155-8165.

Tamames, J. and A. Tramontano (2000). DANTE: a workbench for sequence analysis. *Trends in Biochemical Sciences* **25**(8): 402-403.

Tanaka, K., K. Matsumoto, et al. (1989). Ira1, an Inhibitory Regulator of the Ras-Cyclic Amp Pathway in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* **9**(2): 757-768.

Tatchell, K. (1986). Ras Genes and Growth-Control in *Saccharomyces cerevisiae*. *Journal of Bacteriology* **166**(2): 364-367.

Toda, T., S. Cameron, et al. (1987). Cloning and Characterization of Bcy1, a Locus Encoding a Regulatory Subunit of the Cyclic Amp-Dependent Protein-Kinase in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* **7**(4): 1371-1377.

Toda, T., I. Uno, et al. (1985). In Yeast, Ras Proteins Are Controlling Elements of Adenylate Cyclase. *Cell* **40**(1): 27-36.

Tranque, P., M. C.-Y. Hu, et al. (1998). rRNA complementarity within mRNAs: A possible basis for mRNA- ribosome interactions and translational control. *Proceedings of the National Academy of Sciences of the United States of America* **95**(21): 12238-12243.

van de Peppel, J., P. Kemmeren, et al. (2003). Monitoring global messenger RNA changes in externally controlled microarray experiments. *Embo Reports* **4**(4): 387-393.

van Helden, J., B. André, et al. (2000). A web site for the computational analysis of yeast regulatory sequences. *Yeast* **16**(2): 177-187.

Varani, G. and W. H. McClain (2000). The G·U wobble base pair: A fundamental building block of RNA structure crucial to RNA function in diverse biological systems. *EMBO Reports* **1**(1): 18-23.

Vilo, J., A. Brazma, et al. (2000). Mining for putative regulatory elements in the yeast genome using gene expression data. *ISMB*: 1-10.

Wach, A., A. Brachat, et al. (1996). Guidelines for EUROFAN B0 Program: ORF deletants, plasmid tools, basic functional analyses. **2002**.

Wach, A., A. Brachat, et al. (1994). New Heterologous Modules for Classical or Pcr-Based Gene Disruptions in Saccharomyces-Cerevisiae. *Yeast* **10**(13): 1793-1808.

Walker, G. M. (1998). Yeast Physiology and Biotechnology, *John Wiley and Sons*, ISBN: 0-471-96446-8.

Walsh, S. and B. Barrell (1996). The *Saccharomyces cerevisiae* genome on the World Wide Web. *Trends in Genetics* **12**(7): 276-277.

Wei, W., P. Nurse, et al. (1993). Yeast-Cells Can Enter a Quiescent State through G_1, S, G_2, or M-Phase of the Cell-Cycle. *Cancer Research* **53**(8): 1867-1870.

Werner-Washburne, M., E. Braun, et al. (1993). Stationary-Phase in the Yeast *Saccharomyces cerevisiae*. *Microbiological Reviews* **57**(2): 383-401.

Werner-Washburne, M., E. L. Braun, et al. (1996). Stationary phase in *Saccharomyces cerevisiae*. *Molecular Microbiology* **19**(6): 1159-1166.

Werner-Washburne, M., B. Wylie, et al. (2002). Comparative analysis of multiple genome-scale data sets. *Genome Research* **12**(10): 1564-1573.

Wiebe, M. G. (2003). *Saccharomyces cerevisiae* - Life and cell cycle.

Wilson, W. A., S. A. Hawley, et al. (1996). Glucose repression/derepression in budding yeast: SNF1 protein kinase is activated by phosphorylation under derepressing conditions, and this correlates with a high AMP:ATP ratio. *Current Biology* **6**(11): 1426-1434.

Winston, F., C. Dollard, et al. (1995). Construction of a Set of Convenient *Saccharomyces cerevisiae* Strains That Are Isogenic to S288c. *Yeast* **11**(1): 53-55.

Wittenburg, C. and S. I. Reed (1996). Plugging it in: Signaling circuits and the yeast cell cycle. *Current Opinion in Cell Biology* **8**(2): 223-230.

Wolfsberg, T. G., S. Schafer, et al. (2001). Organelle genome resources at NCBI. *Trends in Biochemical Sciences* **26**(3): 199-203.

Wood, V., K. M. Rutherford, et al. (2001). A re-annotation of the *Saccharomyces cerevisiae* genome. *Comparative and Functional Genomics* **2**(3): 143-154.

Yang, R., K. T. Chun, et al. (1998). Mitochondrial respiratory mutants in yeast inhibit glycogen accumulation by blocking activation of glycogen synthase. *Journal of Biological Chemistry* **273**(47): 31337-31344.

Zaragoza, D., A. Ghavidel, et al. (1998). Rapamycin induces the G_0 program of transcriptional repression in yeast by interfering with the TOR signaling pathway. *Molecular and Cellular Biology* **18**(8): 4463-4470.

Zhou, W., G. M. Edelman, et al. (2001). Transcript leader regions of two *Saccharomyces cerevisiae* mRNAs contain internal ribosome entry sites that function in living cells. *Proceedings of the National Academy of Sciences of the United States of America* **98**(4): 1531-1536.

APPENDIX 1

The results of testing ORF knockouts for stationary phase defects in various different media. Respiratory competency was assumed if the knockout strain was able to grown on YPG medium. Strains with no stationary phase defect on retesting in YPD medium are highlighted. "Yes" and "No" indicate whether an ORF knockout strain does or does not have a stationary phase defect in that medium (a loss of viability after three months in spent medium). "?" and "-" respectively indicate that no conclusion could be drawn or there are no results. Respiratory competency ("resp. comp. Column) is defined as the ability ("Yes") or inability ("No") to grow on YPG medium. Highlighted strains do not show a defect when grown in YPD medium.

ORF	Defect in YPD medium				Defect in SC medium		Other media		
	2% dextrose	0.3% dextrose	With GSH	With Sorbitol	2% dextrose	0.3% dextrose	YPG	YPGal	resp. comp.
YBR179c	Yes	No	No	No	?	No	No	No	Yes
YBR268w	Yes	No	No	No	No	No	No	No	Yes
YCR003w	Yes	Yes	No	No	Yes	Yes	-	?	No
YCR046c	Yes	No	No	Yes	Yes	Yes	-	Yes	No
YDL044c	Yes	No	No	Yes	Yes	Yes	Yes	Yes	Yes
YDL045w-a	Yes	No	Yes	Yes	Yes	Yes	-	Yes	No
YDL067c	No	No	?	No	Yes	No	No	Yes	Yes
YDL068w	Yes	No	No	No	Yes	Yes	-	Yes	No
YDL107w	Yes	No	No	Yes	Yes	Yes	-	Yes	No
YDL202w	Yes	Yes	No	Yes	Yes	Yes	-	?	No
YDR059c	No	No	No	No	No	No	No	No	Yes
YDR065w	Yes	No	No	Yes	Yes	Yes	-	Yes	No
YDR115w	Yes	No	No	Yes	Yes	No	No	Yes	Yes
YDR138w	Yes	Yes	No	No	No	No	No	No	Yes
YDR148c	Yes	No	No	No	?	Yes	-	Yes	No

ORF	Defect in YPD medium				Defect in SC medium		Other media		
	2% dextrose	0.3% dextrose	With GSH	With Sorbitol	2% dextrose	0.3% dextrose	YPG	YPGal	resp. comp.
YDR175c	Yes	No	No	Yes	Yes	No	-	Yes	No
YDR178w	Yes	No	Yes	No	No	No	No	Yes	Yes
YDR194c	Yes	No	No	Yes	Yes	Yes	Yes	Yes	Yes
YDR197w	No	No	No	Yes	Yes	Yes	-	Yes	No
YDR204w	Yes	No	No	?	Yes	Yes	-	No	No
YDR237w	No	No	No	No	Yes	Yes	?	No	Yes
YDR298c	Yes	No	No	No	?	No	No	No	Yes
YDR337w	Yes	?	?	Yes	?	?	?	No	Yes
YDR405w	Yes	No	No	Yes	Yes	No	No	No	Yes
YDR507c	Yes	No	No	?	Yes	No	No	No	Yes
YDR511w	No	No	No	No	Yes	No	No	No	Yes
YDR518w	Yes	No	No	Yes	Yes	Yes	-	Yes	No
YDR523c	Yes	Yes	No	Yes	Yes	No	-	?	No
YER087w	No	No	No	No	No	No	No	No	Yes
YER131w	No	No	No	No	?	No	No	?	Yes
YER154w	Yes	No	No	No	Yes	Yes	-	No	No
YER155c	No	No	No	?	Yes	No	No	Yes	Yes
YFL036w	Yes	No	No	No	?	No	?	No	Yes
YGL107c	Yes	No	No	?	Yes	Yes	-	No	No
YGL129c	Yes	No	No	No	Yes	Yes	-	No	No
YGL135w	Yes	Yes	No	?	Yes	Yes	-	?	No
YGL143c	Yes	No	No	Yes	Yes	Yes	-	Yes	No
YGL246c	No	No	Yes	No	Yes	Yes	No	No	Yes
YGR062c	Yes	No	No	No	Yes	Yes	-	No	No
YGR076c	Yes	No	No	No	Yes	Yes	-	?	No
YGR102c	Yes	No	No	No	Yes	Yes	-	?	No
YGR150c	Yes	No	No	?	Yes	Yes	-	?	No
YGR160w	No	No	No	No	Yes	No	No	?	Yes

ORF	Defect in YPD medium				Defect in SC medium		Other media		
	2% dextrose	0.3% dextrose	With GSH	With Sorbitol	2% dextrose	0.3% dextrose	YPG	YPGal	resp. comp.
YGR171c	Yes	No	No	No	No	No	No	No	Yes
YGR220c	Yes	No	No	No	Yes	Yes	-	No	No
YHL038c	Yes	No	No	?	No	No	No	No	Yes
YHR038w	Yes	No	No	No	No	No	No	No	Yes
YHR120w	Yes	No	No	No	Yes	Yes	-	?	No
YHR147c	Yes	No	No	No	Yes	Yes	-	Yes	No
YIL097w	No	No	?	No	No	No	No	No	Yes
YIR021w	Yes	No	No	Yes	Yes	Yes	-	Yes	No
YJL063c	Yes	No	No	?	Yes	Yes	-	No	No
YJL102w	Yes	No	No	No	Yes	Yes	-	No	No
YJL188c	Yes	Yes	No	?	Yes	Yes	Yes	Yes	Yes
YJR113c	Yes	No	No	Yes	Yes	Yes	-	Yes	No
YJR122w	Yes	No	?	Yes	Yes	Yes	-	Yes	No
YJR144w	Yes	No	No	Yes	Yes	Yes	Yes	Yes	Yes
YKL040c	No	No	?	?	No	No	No	No	Yes
YKL085w	Yes	No	?	?	Yes	Yes	Yes	No	Yes
YKL134c	No	No	?	No	No	No	No	No	Yes
YKL138c	?	No	?	No	Yes	?	No	No	Yes
YKL169c	No	No	?	No	No	No	No	No	Yes
YKR006c	Yes	No	?	Yes	Yes	Yes	Yes	?	Yes
YKR024c	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
YKR085c	Yes	No	No	Yes	Yes	Yes	-	?	No
YKR097w	No	No	No	No	No	No	No	No	Yes
YLL009c	No	No	?	?	No	No	No	No	Yes
YLL018c-a	Yes	No	Yes	Yes	Yes	Yes	-	?	No
YLL027w	No	No	?	No	No	No	No	No	Yes
YLL041c	Yes	?	Yes	Yes	Yes	Yes	-	No	No
YLR067c	Yes	?	No	Yes	Yes	Yes	-	No	No

ORF	Defect in YPD medium				Defect in SC medium		Other media		
	2% dextrose	0.3% dextrose	With GSH	With Sorbitol	2% dextrose	0.3% dextrose	YPG	YPGal	resp. comp.
YLR069c	Yes	No	No	No	Yes	Yes	-	No	No
YLR203c	Yes	No	No	?	Yes	Yes	-	No	No
YLR260w	Yes	No	No	Yes	Yes	Yes	-	Yes	No
YLR295c	No	No	No	No	No	No	No	No	Yes
YLR358c	?	No	No	No	?	Yes	No	No	Yes
YML061c	?	No	No	?	Yes	Yes	Yes	?	Yes
YML110c	Yes	No	No	Yes	Yes	Yes	?	Yes	Yes
YML129c	Yes	No	No	No	Yes	No	No	No	Yes
YMR024w	?	No	No	No	No	No	No	No	Yes
YMR064w	Yes	No	No	Yes	Yes	Yes	Yes	Yes	Yes
YMR098c	?	No	No	No	Yes	Yes	-	No	No
YMR150c	Yes	No	No	No	Yes	Yes	-	No	No
YMR158w	Yes	No	No	Yes	Yes	Yes	-	Yes	No
YMR193w	Yes	No	No	Yes	Yes	Yes	-	Yes	No
YMR228w	Yes	No	No	Yes	Yes	Yes	-	Yes	No
YMR267w	Yes	No	No	No	No	No	No	No	Yes
YMR282c	Yes	No	No	No	No	No	No	No	Yes
YNL005c	No	No	No	No	No	No	No	No	Yes
YNL037c	No	No	No	No	No	No	No	No	Yes
YNL139c	Yes	Yes	No	No	Yes	Yes	Yes	No	Yes
YNL177c	Yes	No	No	No	Yes	No	No	No	Yes
YNL225c	No	No	No	No	Yes	Yes	No	No	Yes
YNL284c	No	No	No	No	No	Yes	No	No	Yes
YNR036c	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
YNR037c	No	No	0	?	Yes	No	No	No	Yes
YNR041c	Yes	No	Yes	Yes	Yes	Yes	-	?	No
YOL012c	No	No	0	Yes	No	No	No	No	Yes
YOL033w	No	No	0	?	No	No	No	No	Yes

ORF	Defect in YPD medium				Defect in SC medium		Other media		
	2% dextrose	0.3% dextrose	With GSH	With Sorbitol	2% dextrose	0.3% dextrose	YPG	YPGal	resp. comp.
YOL071w	No	No	0	?	No	No	-	No	No
YOL095c	Yes	No	0	No	Yes	Yes	-	No	No
YOL143c	Yes	Yes	0	No	Yes	Yes	-	No	No
YOR305w	Yes	No	0	No	Yes	Yes	-	?	No
YOR358W	No	No	0	?	Yes	No	-	?	No
YPL029w	Yes	No	0	No	Yes	Yes	-	No	No
YPL031c	No	No	0	No	Yes	Yes	-	?	No
YPL040c	Yes	No	No	Yes	Yes	Yes	-	Yes	No
YPL045w	No	No	No	Yes	Yes	Yes	-	No	No
YPL059w	Yes	No	?	No	Yes	No	-	No	No
YPL078c	Yes	No	Yes	Yes	Yes	Yes	-	Yes	No
YPL097w	Yes	No	Yes	Yes	Yes	Yes	No	Yes	Yes
YPL104w	Yes	No	No	Yes	?	?	No	Yes	Yes
YPL172c	Yes	No	Yes	Yes	Yes	No	-	?	No
YPL173w	Yes	No	Yes	Yes	Yes	Yes	-	Yes	No
YPL271w	Yes	No	Yes	Yes	Yes	Yes	No	No	Yes
YPR047w	Yes	No	Yes	Yes	Yes	?	-	Yes	No
YPR067w	Yes	No	Yes	Yes	Yes	Yes	-	Yes	No
YPR099C	Yes	No	Yes	Yes	Yes	Yes	-	No	No
YPR100w	Yes	?	Yes	Yes	Yes	Yes	-	No	No
YPR116w	Yes	No	Yes	Yes	Yes	No	-	Yes	No
YPR124w	No	?	No	No	No	No	No	No	Yes
YPR166c	Yes	No	Yes	Yes	Yes	Yes	-	Yes	No

Appendix 2

Retest Of Mitochondrial Ribosome Genes

ORF knockout strain	Repeat 1								Repeat 2							
	Growth of culture after 7 days				Growth of sample after 93 days				Growth of culture after 7 days				Growth of sample after 93 days			
	a	α	Het	Hom	a	α	Het	Hom	a	α	Het	Hom	a	α	Het	Hom
YBL038w	Yes	Yes	Yes	Yes	C	C	IV ~ 50	C	Yes	Yes	Yes	Yes	C	G	G	C
YBR122c	Yes	No	No	No	-	-	-	-	Yes	No	No	No	-	-	-	-
YBR146w	Yes	Yes	Yes	Yes	C	-	C	3	Yes	Yes	Yes	Yes	C	-	33	~ 50
YBR251w	Yes	Yes	Yes	Yes	C	~ 50	~ 50	7	Yes	Yes	Yes	Yes	C	C	16	C
YBR282w	No	No	Yes	No	-	-	C	-	No	No	Yes	No	-	-	-	-
YDR322w	Yes	Yes	Yes	No	C	-	C	-	Yes	Yes	Yes	No	C	-	-	-
YDR347w	Yes	Yes	Yes	No	C	C	G	G	Yes	Yes	Yes	No	C	C	IV ~ 50	C
YDR462w	Yes	Yes	Yes	Yes	C	-	C	-	Yes	Yes	Yes	Yes	G	-	C	-
YFR049w	Yes	Yes	Yes	Yes	C	C	25	C	Yes	Yes	Yes	Yes	C	IV ~ 50	~ 50	C

Key to table: a = MAT**a** strain; α = MATα strain; Het = Heterozygous diploid strain; Hom = Homozygous diploid strain; C = Confluent growth of sample; G = Grainy - growth of sample is beginning to become confluent; *number* = A number of discrete colonies were grown; - = No result available.

ORF knockout strain	Repeat 1								Repeat 2							
	Growth of culture after 7 days			Growth of sample after 93 days					Growth of culture after 7 days				Growth of sample after 93 days			
	a	α	Het	Hom	a	α	Het	Hom	a	α	Het	Hom	a	α	Het	Hom
YGR084c	Yes	Yes	Yes	Yes	C	C	G	≥ 50	Yes	Yes	Yes	Yes	C	C	C	C
YHR075c	Yes	Yes	Yes	Yes	C	C	C	C	Yes	Yes	Yes	Yes	C	C	C	28
YJL096w	Yes	Yes	Yes	Yes	-	-	G	-	Yes	Yes	Yes	Yes	-	-	C	-
YKL003c	Yes	No	Yes	Yes	-	-	C	-	Yes	No	Yes	Yes	-	-	C	-
YKL142w	Yes	No	Yes	Yes	C	-	≥ 50	C	Yes	No	Yes	Yes	C	-	C	G
YKL167c	No	No	No	Yes	-	-	-	≥ 50	No	No	No	Yes	-	-	-	≥ 50
YKL170w	Yes	Yes	Yes	Yes	C	-	C	G	Yes	Yes	Yes	Yes	C	-	G	C
YLR312w-a	Yes	Yes	Yes	Yes	C	C	G	C	Yes	Yes	Yes	Yes	C	G	C	≥ 50
YLR439w	Yes	Yes	Yes	Yes	-	-	C	-	Yes	Yes	Yes	Yes	C	-	C	-

Key to table: a = MAT**a** strain; α = MATα strain; Het = Heterozygous diploid strain; Hom = Homozygous diploid strain; C = Confluent growth of sample; G = Grainy - growth of sample is beginning to become confluent; *number* = A number of discrete colonies were grown; - = No result available.

ORF knockout strain	Repeat 1								Repeat 2							
	Growth of culture after 7 days				Growth of sample after 93 days				Growth of culture after 7 days				Growth of sample after 93 days			
	a	α	Het	Hom	a	α	Het	Hom	a	α	Het	Hom	a	α	Het	Hom
YMR225c	Yes	Yes	Yes	Yes	C	C	C	≥ 50	Yes	Yes	Yes	Yes	C	C	C	G
YMR286w	Yes	Yes	Yes	Yes	C	C	G	G	Yes	Yes	Yes	Yes	C	C	≥ 50	C
YNL137c	No	No	Yes	No	0	0	C	0	No	No	Yes	No	0	0	G	0
YNL185c	No	Yes	Yes	Yes	-	-	C	18	No	Yes	Yes	Yes	-	-	C	37
YNL252c	Yes	Yes	Yes	Yes	-	C	G	-	Yes	Yes	Yes	Yes	-	C	≥ 50	-
YNL306w	Yes	No	Yes	No	C	-	G	-	Yes	No	Yes	No	C	-	G	-
YOR150w	Yes	No	Yes	No	C	0	C	-	Yes	No	Yes	No	C	0	~ 50	-
YOR158w	No	No	Yes	No	-	-	C	-	No	No	Yes	No	-	-	C	-
FY1679	Yes				C				Yes				~ 50			

Key to table: a = MAT**a** strain; α = MATα strain; Het = Heterozygous diploid strain; Hom = Homozygous diploid strain; C = Confluent growth of sample; G = Grainy - growth of sample is beginning to become confluent; *number* = A number of discrete colonies were grown; - = No result available.

Retest Of Respiratory Deficient ORF knockout Strains

Key to tables: A = MAT**a** strain; α = MATα strain; C = Confluent growth of sample; G = Grainy - growth of sample is beginning to become confluent; *number* = A number of discrete colonies were grown; - = No result available; ? = No definite conclusion can be drawn. Gene families are highlighted in grey. The three *SP defect* columns are; "Definite": All strains indicate the same conclusion, "Possible": No result for one or more strains therefore the conclusion could potentially be wrong, "1° Screen": Results of the screen of the entire SGDP set of ORF knockouts.

Gene		93 day growth				SP defect		
		A	A	α	α	Definite	Possible	1° screen
Abf2	YMR072W	-	-			?	No	No
Acn9	YDR511W	C	C	-	-	?	?	Yes
Aco1	YLR304C	C		-	-	?	?	No
Aep1	YMR064W	-	4	-	-	Yes		Yes
Atg17	YLR423C	C	C	C	C	No		No
Atp2	YJR121W	C	C			?	No	No
Atp4	YPL078C	-	1	-	-	Yes		Yes
Atp10	YLR393W	C	C	C	C	No		No
ATP11	YNL315C	C	C	C	-	No		No
Atp12	YJL180C	C	C	-		?	?	No
Atp14	YLR295C	-	-	-	1	Yes		Yes
Atp15	YPL271W	-	2	1	2	Yes		Yes
Atp18	YML081C-A	C	C	C	C	No		No
Atx1	YNL259C	C	C	C	C	No		No
Bro1	YPL084W	C	C	C	C	No		No
Caf17	YJR122W	-	~10	-	-	Yes		Yes
Cbp1	YJL209W				-	?	Yes	No
Cbp3	YPL215W	C	19	-	-	Yes		No
Cbs1	YDL069C	-	-	C	C	?	?	No
Cbs2	YDR197W	-	C	-	-	Yes		Yes
Cbt1	YKL208W	C	C	C	C	No		No
Cin4	YMR138W	C	-	C	-	?	?	No
Cit1	YNR001C	C	C	C	-	No		No
Cog1	YGL223C				1	?	Yes	No
Coq2	YNR041C	-	-	-	-	Yes		Yes
Coq3	YOL096C	C	C	-	-	?	?	No
Coq4	YDR204W	-	-	-	-	Yes		Yes
Coq5	YML110C		C	-		?	?	Yes

Gene		93 day growth				SP defect		
		A	A	α	α	Definite	Possible	1° screen
Coq6	YGR255C	Con	Con		-	?	Yes	No
Cox5a	YNL052W					?	?	No
Cox6	YHR051W			G	-	?	?	No
Cox8	YLR395C	C	C	C	C	No		No
Cox9	YDL067C	1	-	-	-	Yes		Yes
Cox10	YPL172C	-	-	3	-	Yes		Yes
Cox11	YPL132W	-	-		-	Yes		No
Cox12	YLR038C	8	-	C	C	?	?	No
Cox14	YML129C	-	-	1		Yes		Yes
Cox15	YER141W		-	G		?	?	No
Cox16	YJL003W	1	1			?	Yes	No
Cox18	YGR062C	-		-	-	Yes		Yes
Cox19	YLL018C-A		-	-	-	Yes		Yes
Cox20	YDR231C	-	-	C	C	?	?	No
Crd1	YDL142C	C	C	C	C	No		No
Csf1	YLR087C	C	C	C	Con	No		No
Ctr1	YPR124W	~50			-	Yes		Yes
Cwh36	YCL007C			-	-	?	Yes	No
Cyc1	YJR048W	C	C	C	C	No		No
Dcs1	YLR270W	-	-	C	C	?	?	No
Dem1	YBR163W	-	-	C	C	?	?	No
Dip5	YPL265W	C	C	C	C	No		No
Doc1	YGL240W					?	?	No
Eaf7	YNL136W	C	C	C	C	No		No
Emi1	YDR512C	Con	G	C	C	No		No
Emp70	YLR083C	C	C	C	C	No		No
Erg2	YMR202W	C	C	C	C	No		No
Erg3	YLR056W	C	G	3	29	?	?	No
Erg24	YNL280C	C	C	C	C	No		No
Eug1	YDR518W	-	-	-	-	Yes		Yes
Fah1	YDL045C	C	Con	Con	Con	?	Yes	No
Fbp1	YLR377C	Con	Con	Con	~50/Con	?	?	No
Fbp26	YJL155C	C	C	C	C	No		No
Fet3	YMR058W		C	C	-	?	No	No
Fmc1	YIL098C	-	C			?	No	No
Fmp13	YKR016W	C	G	1	C	No		No
Fmp21	YBR269C	-		C	C	?	?	No
Fmp25	YLR077W	C	C	C	C	No		No
Fmp35	YIL157C	C	C	C	C	No		No
Fmp36	YDR493W	Con	Con	Con	C	?	No	No
Fmp38	YOR205C			Con	Con	?	?	No
Fpr2	YDR519W	C	C	C	C	No		No
Ftr1	YER145C	C	C		-	?	No	No

Gene		93 day growth				SP defect		
		A	A	α	α	Definite	Possible	1° screen
Fum1	YPL262W	C	C	C	C	No		No
Gef1	YJR040W	C	C	C	C	No		No
Gin4	YDR507C	-	-	1	-	Yes		Yes
Glo3	YER122C	-				?	Yes	No
Gon5	YPL183W-A		Con	Con	Con	?	?	No
Grx3	YDR098C	-	-		-	Yes		No
Hap2	YGL237C					?	?	No
Hap5	YOR358W	-	-	-	6	Yes		Yes
Hcm1	YCR065W	C	C	C	C	No		No
Hfa1	YMR207C	C	C	C	C	No		No
Hfi1	YPL254W	C	C			?	No	No
Hmi1	YOL095C	1	-	-	-	Yes		Yes
Hta1	YDR225W	C	Con	C	C	No		No
Idh2	YOR136W	G	C	G	C	No		No
Img2	YCR071C			-		?	Yes	No
Imp1	YMR150C					?	?	Yes
Isa2	YPR067W		-			?	Yes	Yes
Ism1	YPL040C	1	-	-	-	Yes		Yes
Kap123	YER110C					?	?	No
Kgd1	YIL125W	C	C	C	C	No		No
Kgd2	YDR148C	G	-	21	C	?	?	Yes
Kha1	YJL094C	C	C	Con	Con	?	No	No
Lat1	YNL071W					?	?	No
Lcb5	YLR260W	-	-	-	-	Yes		Yes
Lip2	YLR239C	≥ 50	G			?	No	No
Lip5	YOR196C	C	C			?	No	No
Lpd1	YFL018C	C	-	C	-	?	?	No
Lpe10	YPL060W	C	C	C	C	No		No
Mam33	YIL070C		Con	C	C	?	No	No
Mbp1	YDL056W	C	C	-	-	?	?	No
Mct1	YOR221C	C	C	-	-	?	?	No
Mdj1	YFL016C	Con	Con	Con	C	?	No	No
Mdl2	YPL270W	C	C	C	C	No		No
Mdm30	YLR368W	C	C	C	C	No		No
Mdm35	YKL053C-A	-	C	C	C	No		No
Mef1	YLR069C	-	-	-	26	Yes		Yes
Mef2	YJL102W	-	-	-	-	Yes		Yes
Met7	YOR241W	C	-	-		?	Yes	No
Met18	YIL128W	C	C			?	No	No
Mgm1	YOR211C	Con	-	C	C	?	?	No
Mgm101	YJR144W	-	-	27	-	Yes		Yes
Mip1	YOR330C	C	C	-	C	No		No
Mir1	YJR077C	C	C	C	C	No		No

Gene		93 day growth				SP defect		
		A	A	α	α	Definite	Possible	1° screen
Mis1	YBR084W	C	C	C	C	No		No
Mlh1	YMR167W	C	C	C	C	No		No
Mls1	YNL117W	C	C	C	C	No		No
Mne1	YOR350C	C	C	C	C	No		No
Mog1	YJR074W	C	C	C	C	No		No
Mon2	YNL297C	C	C	C	C	No		No
Mrf1	YGL143C	Con	Con	-		?	Yes	Yes
Mrp2	YPR166C	-	-		-	Yes		Yes
Mrp7	YNL005C	-	-	-	3	Yes		Yes
Mrp13	YGR084C	C	Con	C	C	No		No
Mrpl4	YLR439W					?	?	No
Mrpl6	YHR147C	-	1	2	-	Yes		Yes
Mrpl7	YDR237W	-	-	-	-	Yes		Yes
Mrpl8	YJL063C	-	-	2	-	Yes		Yes
Mrpl9	YGR220C	1	1	1	-	Yes		Yes
Mrpl10	YNL284C		Con	Con	Con	?	?	Yes
Mrpl15	YLR312W-A	C	C	C	C	No		No
Mrpl17	YNL252C	1		C	C	?	?	No
Mrpl20	YKR085C	-		-	-	Yes		Yes
Mrpl24	YMR193W	-	-	1	-	Yes		Yes
Mrpl25	YGR076C	-	-	-	16	Yes		Yes
Mrpl27	YBR282W					?	?	No
Mrpl28	YDR462W	C	C	-	2	?	?	No
Mrpl32	YCR003W	-	-			?	Yes	Yes
Mrpl40	YPL173W	6	C	-	-	Yes		Yes
Mrpl49	YJL096W	-	-	-		Yes		No
Mrpl51	YPR100W		1		-	Yes		Yes
Mrps1	YGR220C	C	C	C	C	No		Yes
Mrps8	YMR158W	-	-	-	-	Yes		Yes
Mrps16	YPL013C		-	C	C	?	?	No
Mrps17	YMR188C	C	C	-	-	?	?	No
Mrs1	YIR021W	-	-	-	-	Yes		Yes
Mrs2	YOR334W	C	C	C	C	No		No
Msc6	YOR354C	C	C	C	C	No		No
Msd1	YPL104W	1	3	-	-	Yes		Yes
Msf1	YPR047W	-	1	1	-	Yes		Yes
Msh1	YHR120W	-	-	-	-	Yes		Yes
Msk1	YNL073W				-	?	Yes	No
Msl1	YIR009W	C	C	C	C	No		No
Msr1	YHR091C	-	-	-	-	Yes		No
Mss2	YDL107W	-	-	1	-	Yes		Yes
Mss18	YPR134W	C	C	27	C	No		No
Mss51	YLR203C	1	-	C	-	Yes		Yes

Gene		93 day growth				SP defect		
		A	A	α	α	Definite	Possible	1° screen
Mss116	YDR194C	-	-	-	-	Yes		Yes
Msy1	YPL097W	-	-	9	15	Yes		Yes
Mtf1	YMR228W	-	-			?	Yes	Yes
Mtf2	YDL044C	-	-	-	-	Yes		Yes
Mtg1	YMR097C	C	C	C	C	No		No
Mtm1	YGR257C	Con	-	C	C	?	?	No
Mto1	YGL236C					?	?	No
Nam2	YLR382C				-	?	Yes	No
Nbp2	YDR162C	C	C	C	C	No		No
Ndi1	YML120C	C	C	C	C	No		No
Nhx1	YDR456W	C	C	C	C	No		No
Nip100	YPL174C	C	C	C	C	No		No
Num1	YDR150W	C	C	C	C	No		No
Oxa1	YER154W					?	?	Yes
Oye2	YHR179W	C	C	C	C	No		No
Pcp1	YGR101W	C	C	C	C	No		No
Pda1	YER178W			1	1	?	Yes	No
Pde2	YOR360C	C	C	C	C	No		No
Per1	YCR044C					?	?	No
Pet8	YNL003C	-	3		C	?	?	No
Pet54	YGR222W	G	-	C	C	No		No
Pet56	YOR201C	Con	Con	Con	Con	?	?	No
Pet122	YER153C			-	-	?	Yes	No
Pet130	YJL023C	C	C	C	C	No		No
Pet494	YNR045W	C	C	-	G	No		No
Pho2	YDL106C	C	C	C	C	No		No
Pho4	YFR034C	C	C	C	C/Con	No		No
Pho23	YNL097C	-		-		No		No
Pho85	YPL031C	-	-	-	C	Yes		Yes
Pho86	YJL117W	C	C	C	C	No		No
Pif1	YML061C	Con	Con	0	1	?	Yes	Yes
Pkr1	YMR123W	C	C	C	C	No		No
Pmr1	YGL167C	Con	Con		-	?	Yes	No
Pos5	YPL188W	C	C	C	C	No		No
Ppa1	YHR026W	C	C			?	No	No
Ppt2	YPL148C		Con	Con	Con	?	?	No
Ptc1	YDL006W			C	C	?	No	No
Pth1	YHR189W	C	C	C	C	No		No
Qcr2	YPR191W	-	-	-	-	Yes		No
Qcr6	YFR033C	C	C	C	C	No		No
Qcr7	YDR529C	-	-	-	-	Yes		No
Qcr8	YJL166W	C	C	C	C	No		No
Qri5	YLR204W	-	-	C	C	?	?	No

Gene		93 day growth				SP defect		
		A	A	α	α	Definite	Possible	1° screen
Qri7	YDL104C	-	1	C	C	?	?	No
Rav1	YJR033C	C	C	C	C	No		No
Ref2	YDR195W	C	C	C	C	No		No
Rex2	YLR059C	C	C	C	C	No		No
Rim1	YCR028C-A	-	-			?	?	Yes
Rim9	YMR063W	C	C	C	C	No		No
Rmd9	YGL107C	-	1	2	≥ 50	Yes		Yes
Rmd12	YHR067W	C	C	2	G	No		No
Rpe1	YJL121C	C	C	C	C	No		No
Rpl1b	YGL135W	Con	Con		1	?	Yes	Yes
Rpo41	YFL036W			21		?	Yes	Yes
Rsm7	YJR113C	3	1	-	-	Yes		Yes
Rsm19	YNR037C	-	-	-	-	Yes		Yes
Rsm23	YGL129C	Con		1	-	?	Yes	Yes
Rsm24	YDR175C	-	-	-	-	Yes		Yes
Rsm28	YDR494W	C	C	C	C	No		No
Rvs161	YCR009C	C		C	Con	No		No
Sdh4	YDR178W	1	-	-	-	Yes		Yes
Sec22	YLR268W	C	C	C	C	No		No
Shy1	YGR112W	-	-		-	Yes		No
Smi1	YGR229C	C	C	C	C	No		No
Snf1	YDR477W	~50	~50	~40	~40	No		No
Snf2	YOR290C	-	-		-	Yes		No
Snf4	YGL115W	C	≥ 50	C	C	No		No
Snf5	YBR289W					?	?	No
Snf6	YHL025W	-				?	?	No
Snf8	YPL002C	C	C	C	C	No		No
Sod2	YHR008C	C	C	11	6	?	?	No
Sov1	YMR066W	-	-	C	C	?	?	No
Sps2	YDR522C	C	C	C	C	No		No
Srb8	YCR081W			C	C	?	No	No
Srf1	YGL218W					?	?	No
Ssn3	YPL042C	G	C	C	C	No		No
Ssq1	YLR369W	1	-	Con	-	Yes		No
Stp22	YCL008C	G	C	C	C	No		No
Stv1	YMR054W	-	C	C	C	No		No
Suv3	YPL029W	-	-	-	-	Yes		Yes
Sws2	YNL081C	C	C			?	No	No
Sys1	YJL004C	C	C	C	C	No		No
Tfp1	YDL185W	C	C	C	C	No		No
Tfp3	YPL234C		~50	C	C	No		No
Thr1	YHR025W	C	C	C	C	No		No
Tom5	YPR133W-A	Con	Con	Con	Con	?	?	No

Gene		93 day growth				SP defect		
		A	A	α	α	Definite	Possible	1° screen
Tuf1	YOR187W	-				?	Yes	No
Tus1	YLR425W	C	C	C	C	No		No
Uba3	YPR066W	C	C	2	-	?	?	No
Ubc11	YOR339C	C	C	C	C	No		No
Uma4	YIL031W			C	C	?	No	No
Uma13	YIL031W	C	-		4	?	Yes	No
Uth1	YKR042W	C	C	C	C	No		No
Vma2	YBR127C	C	-	C	C	No		No
Vma6	YLR447C	-	C		C	?	No	No
Vma21	YGR105W	C	2			?	?	No
Vps1	YKR001C	C	C	C	C	No		No
Vps3	YDR495C	C	C	C	C	No		No
Vps4	YPR173C	C	Con	C	C	No		No
Vps20	YMR077C	C	C			?	No	No
Vps25	YJR102C	C	C	G	C	No		No
Vps27	YNR006W	C	C	C	C	No		No
Vps28	YPL065W	C	C	C	C	No		No
Vps36	YLR417W	Con	Con	C	C	?	No	No
Vps52	YDR484W	C	C	C	C	No		No
Ydj1	YNL064C					?	?	No
Yim2	YMR151W	-	-	23	31	Yes		No
Yps16	YLR120C			G	C	?	No	No
Yps33	YLR121C	-	4	-	Con	Yes		No
Yps53	YGL259W			C	C	?	No	No
Ypt6	YLR262C	C	C	C	C	No		No
Yta12	YMR089C	-	-	C	C	?	?	No
YCR024c		-				?	Yes	No
YCR091w		C	C	-	-	?	?	No
YDL032w		C	C	C	C	No		No
YDL033c		C	C	C	C	No		No
YDL068w		-	-	-	1	Yes		No
YDL118w		C	C	C	C	No		No
YDL119c		C	C	C	C	No		No
YDR230w			-	-	-	Yes		No
YDR455c		C	C	C	C	No		No
YDR509w		C	C	C	C	No		No
YDR533c		C	C	C	C	No		No
YER077c		C	C	-	-	?	?	No
YER093c		C	C	C	C	No		No
YGL250w				C	C	?	No	No
YGR102c		-	-	-	≥50	Yes		No
YGR219w		2	-	Con	C	?	?	No

Gene		93 day growth				SP defect		
		A	A	α	α	Definite	Possible	1° screen
YHR009c		C	C			?	No	No
YHR100c		C	C	C	C	No		No
YHR116w		C	C	≥50	C	No		No
YHR168w		-	≥ 50	C	25	?	?	No
YIR024c		C	C	C	C	No		No
YJL022w		C	C	C	C	No		No
YJL046w		C	C	C	C	No		No
YJL193w		C	C	C	C	No		No
YJR018w				C	C	?	No	No
YJR079w		C	C	C	C	No		No
YJR120w		C	3	C	C	No		No
YLR218c		C	C	C	C	No		No
YLR235c		C	C	C	C	No		No
YLR294c		C	C	-	-	?	?	No
YLR338w		Con	Con	-	Con	?	Yes	No
YML089c		-	C	C	Con	?	No	No
YML090w		C	C	C	C	No		No
YMR098c		1	1	≥50	-	Yes		No
YNL080c		-		G	C	?	?	No
YNR018w		C	C	C	C	No		No
YNR020c		C	-	C	C	No		No
YNR025c		C	C	C	C	No		No
YNR036c		-	-	-	-	Yes		No
YNR042w		C	C	1	1	?	?	No
YOR135c		C	C	C	-	No		No
YOR199w		C	G		C	No		No
YOR200w		C	C	C	C	No		No
YOR305w		-	-	-	-	Yes		No
YOR331c				-	C	?	?	No
YOR333c		C	C	C	C	No		No
YPL005w		2	-	C	C	?	?	No
YPL066w		C	Con	C	C	No		No
YPL098c		C	C	C	C	No		No
YPL158c		Con	Con	Con	Con	?	?	No
YPR099c			G	-	-	?	?	No
YPR116w		Con	1	-		?	?	No
YPR123c		-		C	C	?	?	No
YPR146c		C	C	C	C	No		No
YPR197c		C	C	C	C	No		No
Total numbers of …					Yes	69	26	67
					No	133	27	257
					?	122	69	0

Appendix 3

Stationary phase essential genes (after primary screen)

EUROFAN2 set of ORF knockouts

YBR284w	YGR262c	YLL051c	YNL224c
YDL115c	YJL006c	YLR015w	YNR019w
YDL120w	YJL020c	YLR018c	YNR047w
YDR065w	YJL100w	YNL019c	YNR049c
YGL085w	YJL193w	YNL021w	YOL036w
YGL099w	YJL217w	YNL022c	YOL125w
YGL129c	YJR070c	YNL072w	YOL268c
YGL179c	YJR080c	YNL119w	
YGR216c	YLL030c	YNL177c	

SGDP set of ORF knockouts

YBR179c	YDR115w	YDR507c	YGL143c
YBR268w	YDR138w	YDR511w	YGL246c
YCR003w	YDR148c	YDR518w	YGR062c
YCR046c	YDR175c	YDR523c	YGR076c
YDL044c	YDR178w	YER087w	YGR102c
YDL045w-a	YDR194c	YER131w	YGR150c
YDL067c	YDR197w	YER154w	YGR160w
YDL068w	YDR204w	YER155c	YGR171c
YDL107w	YDR237w	YFL036w	YGR220c
YDL202w	YDR298c	YGL107c	YHL038c
YDR059c	YDR337w	YGL129c	YHR038w
YDR065w	YDR405w	YGL135w	YHR120w

YHR147c	YLL018c-a	YMR267w	YPL031c
YIL097w	YLL027w	YMR282c	YPL040c
YIR021w	YLL041c	YNL005c	YPL045w
YJL063c	YLR067c	YNL037c	YPL059w
YJL102w	YLR069c	YNL139c	YPL078c
YJL188c	YLR203c	YNL177c	YPL097w
YJR113c	YLR260w	YNL225c	YPL104w
YJR122w	YLR295c	YNL284c	YPL172c
YJR144w	YLR358c	YNR036c	YPL173w
YKL040c	YML061c	YNR037c	YPL271w
YKL085w	YML110c	YNR041c	YPR047w
YKL134c	YML129c	YOL012c	YPR067w
YKL138c	YMR024w	YOL033w	YPR099C
YKL169c	YMR064w	YOL071w	YPR100w
YKR006c	YMR098c	YOL095c	YPR116w
YKR024c	YMR150c	YOL143c	YPR124w
YKR085c	YMR158w	YOR305w	YPR166c
YKR097w	YMR193w	YOR358w	
YLL009c	YMR228w	YPL029w	

Stationary phase essential genes (after secondary screen; SGDP set of ORF knockouts)

Stationary phase defect when grown in YPD medium

YBR179c	YCR003w	YDL044c	YDL068w
YBR268w	YCR046c	YDL045w-a	YDL107w

YDL202w	YGR062c	YKR085c	YNL177c
YDR065w	YGR076c	YLL018c-a	YNR036c
YDR115w	YGR102c	YLL041c	YNR041c
YDR138w	YGR150c	YLR067c	YOL095c
YDR148c	YGR171c	YLR069c	YOL143c
YDR175c	YGR220c	YLR203c	YOR305w
YDR178w	YHL038c	YLR260w	YPL029w
YDR194c	YHR038w	YLR358c	YPL040c
YDR204w	YHR120w	YML061c	YPL059w
YDR298c	YHR147c	YML110c	YPL078c
YDR337w	YIR021w	YML129c	YPL097w
YDR405w	YJL063c	YMR024w	YPL104w
YDR507c	YJL102w	YMR064w	YPL172c
YDR518w	YJL188c	YMR098c	YPL173w
YDR523c	YJR113c	YMR150c	YPL271w
YER154w	YJR122w	YMR158w	YPR047w
YFL036w	YJR144w	YMR193w	YPR067w
YGL107c	YKL085w	YMR228w	YPR099c
YGL129c	YKL138c	YMR267w	YPR100w
YGL135w	YKR006c	YMR282c	YPR116w
YGL143c	YKR024c	YNL139c	YPR166c

Stationary phase defect when grown in SC medium, but not YPD medium

YDL067c	YDR237w	YER155c	YGR160w
YDR197w	YDR511w	YGL246c	YNL225c

YNL284c	YOR358W	YPL045w
YNR037c	YPL031c	

Mitochondrial ribosomal genes not identified in the primary screen for stationary phase essential genes

Genes tested

YBL038w	YDR462w	YKL167c	YNL185c
YBR122c	YFR049w	YKL170w	YNL252c
YBR146w	YGR084c	YLR312w-a	YNL306w
YBR251w	YHR075c	YLR439w	YOR150w
YBR282w	YJL096w	YMR225c	YOR158w
YDR322w	YKL003c	YMR286w	
YDR347w	YKL142w	YNL137c	

Genes not tested

Q0140 Gene on the mitochondrial genome and not part of the SGDP or EUROFAN set of ORF knockouts.

YHL004w ORF knockout not part of the SGDP or EUROFAN set of ORF knockouts.

Stationary phase essential genes used during the investigation of IRES

YBR179c	YDL045w-a	YDR065w	YDR178w
YBR268w	YDL067c	YDR115w	YDR194c
YCR003w	YDL068w	YDR138w	YDR197w
YCR046c	YDL107w	YDR148c	YDR204w
YDL044c	YDL202w	YDR175c	YDR237w

YDR298c	YHR038w	YML061c	YPL031c
YDR337w	YHR120w	YML110c	YPL040c
YDR405w	YHR147c	YML129c	YPL045w
YDR507c	YIR021w	YMR064w	YPL059w
YDR518w	YJL063c	YMR098c	YPL059w
YDR523c	YJL102w	YMR150c	YPL078c
YER154w	YJL188c	YMR158w	YPL097w
YFL036w	YJR113c	YMR193w	YPL097w
YGL107c	YJR122w	YMR228w	YPL104w
YGL129c	YJR144w	YMR267w	YPL172c
YGL135w	YKL085w	YMR282c	YPL173w
YGL143c	YKL138c	YNL139c	YPL271w
YGL246c	YKR006c	YNL177c	YPR047w
YGR062c	YKR024c	YNL225c	YPR067w
YGR076c	YKR085c	YNR036c	YPR099c
YGR102c	YLL018c-a	YNR037c	YPR100w
YGR150c	YLL041c	YNR041c	YPR116w
YGR160w	YLR067c	YOL095c	YPR166c
YGR171c	YLR069c	YOL143c	
YGR220c	YLR203c	YOR305w	
YHL038c	YLR260w	YPL029w	

APPENDIX 4

Results of loss of viability experiment

The tables in this appendix show the number of viable cells in a number of different cultures over a period of 13 weeks. The number of viable cells was calculated from the number of colonies grown from a 200 µl sample of a dilution of each of the cultures. Cultures were grown of *S. cerevisiae* FY1679 (as a control) and of MAT**a** and MATα haploid and heterozygous and homozygous diploid strains of nine different ORF knockouts. The results are show to 2 decimal places.

***S. cerevisiae* FY1679 cultures**

Week	1	2	3	4	5	6	7
Washed	no data	$1.15x10^8$	$9.80x10^7$	$5.30x10^7$	$6.45x10^7$	$2.85x10^7$	$7.50x10^6$
Unwashed	$7.90x10^7$	$1.04x10^8$	$1.41x10^8$	$8.15x10^7$	$4.75x10^7$	$2.30x10^7$	$2.50x10^6$

Week	8	9	10	11	12	13
Washed	$6.05x10^6$	$5.00x10^7$	$3.80x10^6$	$4.45x10^6$	$5.15x10^6$	$5.50x10^6$
Unwashed	$3.70x10^6$	$2.40x10^7$	$1.30x10^6$	$2.50x10^6$	$2.00x10^6$	$8.50x10^5$

MATa, unwashed cultures

Week	1	2	3	4	5	6	7	8	9	10	11	12	13
YCR003w	$6.35x10^{7}$	$8.65x10^{7}$	$2.00x10^{6}$	$8.50x10^{5}$	$5.00x10^{5}$	$1.00x10^{5}$	$4.03x10^{4}$	no data	no data	$3.00x10^{2}$	$4.50x10^{1}$	$2.00x10^{1}$	no data
YDL202w	$4.75x10^{8}$	$1.14x10^{8}$	$4.90x10^{7}$	no data	$2.50x10^{6}$	$2.15x10^{6}$	$2.50x10^{5}$	$6.50x10^{4}$	$2.50x10^{4}$	$5.00x10^{3}$	no data	no data	no data
YDL405w	$5.45x10^{7}$	no data	$1.90x10^{7}$	$2.15x10^{7}$	$1.80x10^{7}$	$8.93x10^{6}$	$1.18x10^{7}$	$7.30x10^{6}$	$7.90x10^{6}$	$6.00x10^{6}$	$3.85x10^{6}$	$3.00x10^{6}$	$3.45x10^{6}$
YDR194c	no data	no data	no data	no data	no data	no data	no data	no data	no data	no data	no data	no data	no data
YGL135c	$5.40x10^{7}$	$4.05x10^{7}$	$4.25x10^{7}$	$1.80x10^{7}$	$1.80x10^{7}$	no data	no data	no data	no data	no data	no data	no data	no data
YGR220c	$1.33x10^{8}$	$1.07x10^{8}$	$7.25x10^{7}$	$4.05x10^{7}$	$3.35x10^{7}$	$2.95x10^{7}$	$2.00x10^{7}$	$8.00x10^{6}$	$8.65x10^{6}$	$7.35x10^{6}$	$8.80x10^{6}$	no data	$8.00x10^{6}$
YJR113c	$1.27x10^{8}$	$9.65x10^{7}$	$2.50x10^{7}$	$1.88x10^{7}$	$1.30x10^{7}$	$5.05x10^{6}$	$9.50x10^{5}$	$1.30x10^{5}$	$2.50x10^{4}$	$6.00x10^{3}$	no data	$9.35x10^{2}$	no data
YKR085c	$1.49x10^{8}$	$4.50x10^{7}$	$2.30x10^{7}$	$1.22x10^{7}$	$7.00x10^{6}$	$2.20x10^{6}$	$9.00x10^{5}$	$1.00x10^{5}$	$8.50x10^{4}$	$6.00x10^{4}$	$4.00x10^{3}$	$4.80x10^{3}$	$8.00x10^{2}$
YNL177c	$1.45x10^{8}$	$7.15x10^{7}$	$1.80x10^{7}$	$9.50x10^{6}$	$1.10x10^{7}$	$1.33x10^{8}$	$1.50x10^{6}$	$1.00x10^{5}$	no data	no data	no data	no data	$1.10x10^{2}$

MATa, washed cultures

Week	1	2	3	4	5	6	7	8	9	10	11	12	13
YCR003w	$5.85x10^{7}$	$1.00x10^{6}$	$2.92x10^{7}$	$2.00x10^{7}$	$5.00x10^{5}$	no data	no data	5.00E+00	no data	$2.17x10^{3}$	no data	no data	no data
YDL202w	$1.28x10^{8}$	$2.55x10^{7}$	$2.00x10^{6}$	no data	no data	$5.00x10^{4}$	no data	no data	no data	no data	no data	$4.50x10^{1}$	no data
YDL405w	$4.50x10^{7}$	$4.45x10^{7}$	$4.25x10^{7}$	$5.00x10^{6}$	$4.00x10^{7}$	no data	no data	no data	$8.40x10^{6}$	no data	no data	no data	no data
YDR194c	no data	no data	no data	no data	no data	no data	no data	no data	no data	no data	no data	no data	no data
YGL135c	$4.45x10^{7}$	$4.65x10^{7}$	$4.55x10^{7}$	$1.15x10^{7}$	no data	no data	$5.00x10^{5}$	no data	no data	no data	no data	$4.00x10^{5}$	no data
YGR220c	$1.06x10^{8}$	$8.15x10^{7}$	$6.15x10^{7}$	$5.15x10^{7}$	$2.80x10^{7}$	$1.35x10^{7}$	$3.00x10^{6}$	$2.30x10^{6}$	$2.20x10^{6}$	$4.20x10^{6}$	$2.80x10^{6}$	$6.90x10^{6}$	$5.10x10^{6}$
YJR113c	$1.55x10^{8}$	$1.55x10^{7}$	no data	$1.57x10^{5}$	$3.00x10^{4}$	$2.50x10^{3}$	$4.35x10^{2}$	no data	no data	no data	no data	no data	no data
YKR085c	$1.49x10^{8}$	no data	no data	no data	no data	$5.00x10^{1}$	no data	no data	$1.00x10^{1}$	no data	no data	no data	no data
YNL177c	$1.72x10^{8}$	$2.55x10^{7}$	$1.50x10^{6}$	$5.00x10^{5}$	$2.50x10^{6}$	$5.00x10^{5}$	$7.50x10^{5}$	$4.90x10^{5}$	$4.45x10^{5}$	$1.35x10^{5}$	$2.40x10^{5}$	$1.55x10^{5}$	no data

MATα, unwashed cultures

Week	1	2	3	4	5	6	7	8	9	10	11	12	13
YCR003w	$6.35x10^7$	$7.20x10^7$	$4.55x10^7$	$1.75x10^7$	$2.20x10^7$	$1.91x10^7$	$1.80x10^7$	$1.04x10^7$	$1.06x10^7$	$9.35x10^6$	$9.50x10^6$	$7.15x10^6$	$5.00x10^5$
YDL202w	$2.72x10^8$	$7.55x10^7$	$1.45x10^7$	no data	no data	$6.50x10^5$	$1.08x10^5$	$4.50x10^4$	$8.50x10^3$	$4.50x10^3$	no data	no data	no data
YDL405w	$5.35x10^7$	no data	$2.90x10^7$	$1.90x10^7$	$2.25x10^7$	$1.33x10^7$	$6.70x10^6$	$3.75x10^6$	no data	$5.40x10^6$	$3.10x10^6$	$3.80x10^6$	no data
YDR194c	no data	5.00E+00	no data	no data	no data	no data	no data	no data	no data	no data	no data	no data	no data
YGL135c	no data	$1.58x10^8$	$8.50x10^7$	no data	no data	no data	no data	no data	no data	no data	no data	no data	no data
YGR220c	2.28x1	$1.06x10^8$	$8.90x10^7$	$5.45x10^7$	$4.15x10^7$	$2.50x10^7$	$2.00x10^7$	$1.05x10^7$	$8.95x10^6$	$8.50x10^6$	$8.55x10^6$	$9.00x10^6$	$1.00x10^7$
YJR113c	$6.45x10^7$	$6.30x10^7$	$3.30x10^7$	$8.68x10^6$	$2.00x10^6$	$5.50x10^5$	$3.50x10^5$	no data	$1.00x10^4$	no data	$1.45x10^2$	$4.50x10^1$	no data
YKR085c	$1.11x10^8$	no data	$1.65x10^7$	$1.25x10^7$	$6.00x10^6$	$2.20x10^6$	$8.00x10^5$	$2.50x10^6$	$3.50x10^4$	$4.00x10^3$	no data	$6.50x10^2$	$2.60x10^2$
YNL177c	$1.82x10^8$	$4.75x10^7$	no data	no data	$2.75x10^7$	$1.50x10^6$	$4.95x10^6$	$1.85x10^6$	no data	no data	$2.47x10^6$	$2.54x10^6$	$3.30x10^6$

MATα, washed cultures

Week	1	2	3	4	5	6	7	8	9	10	11	12	13
YCR003w	$3.76x10^7$	$3.45x10^7$	$1.95x10^7$	$8.50x10^6$	$1.65x10^7$	$6.30x10^6$	$3.55x10^6$	$1.55x10^6$	$1.57x10^6$	$6.00x10^5$	$5.65x10^5$	no data	$1.40x10^5$
YDL202w	$1.41x10^8$	$1.70x10^7$	no data	no data	no data	no data	no data	no data	no data	no data	no data	$4.00x10^1$	no data
YDL405w	$5.70x10^7$	$3.70x10^7$	$2.95x10^7$	$2.80x10^7$	$7.50x10^6$	$4.15x10^6$	$2.65x10^6$	$2.60x10^6$	$1.85x10^6$	$1.85x10^6$	$1.15x10^6$	$1.25x10^6$	$7.00x10^5$
YDR194c	no data	no data	no data	no data	no data	no data	no data	no data	no data	no data	no data	no data	no data
YGL135c	no data	$3.20x10^7$	$8.00x10^6$	no data	$1.00x10^7$	no data	no data	no data	no data	no data	no data	no data	$3.00x10^5$
YGR220c	$1.18x10^8$	$9.65x10^7$	$5.45x10^7$	$6.55x10^7$	$3.65x10^7$	$2.70x10^7$	$8.00x10^6$	$6.95x10^6$	$5.00x10^5$	$4.60x10^6$	$7.15x10^6$	$5.75x10^6$	$5.35x10^6$
YJR113c	$1.00x10^8$	$2.50x10^6$	$5.00x10^5$	$6.30x10^4$	no data	no data	no data	no data	no data	$1.00x10^1$	no data	no data	no data
YKR085c	$9.20x10^7$	no data	$1.00x10^6$	no data	no data	$7.50x10^2$	$1.05x10^2$	$1.00x10^1$	no data	no data	$8.00x10^1$	no data	no data
YNL177c	$1.89x10^8$	$1.20x10^7$	no data	$1.25x10^7$	no data	no data	no data	no data	no data	no data	no data	no data	no data

Heterozygous diploid, unwashed cultures

Week	1	2	3	4	5	6	7	8	9	10	11	12	13
YCR003w	$2.25x10^{8}$	no data	no data	$5.00x10^{5}$	no data	$7.25x10^{6}$	no data	no data	$5.00x10^{2}$	5.00E+00	no data	no data	no data
YDL202w	$5.20x10^{7}$	$2.75x10^{7}$	no data	no data	$1.80x10^{7}$	$9.55x10^{6}$	$9.00x10^{6}$	$1.12x10^{7}$	$7.35x10^{6}$	$6.10x10^{6}$	$8.05x10^{6}$	$7.65x10^{6}$	$1.50x10^{5}$
YDL405w	$4.55x10^{7}$	$4.90x10^{7}$	$3.95x10^{7}$	$2.90x10^{7}$	$2.25x10^{7}$	$1.78x10^{7}$	$1.28x10^{7}$	$8.30x10^{6}$	$7.65x10^{6}$	$1.11x10^{7}$	$9.60x10^{6}$	$9.25x10^{6}$	$7.30x10^{6}$
YDR194c	$1.11x10^{3}$	$7.80x10^{2}$	no data	$5.45x10^{2}$	$4.15x10^{2}$	$3.70x10^{2}$	$2.40x10^{2}$	no data	$5.50x10^{1}$	$7.10x10^{2}$	$6.55x10^{2}$	$6.30x10^{2}$	$7.95x10^{2}$
YGL135c	$4.80x10^{7}$	$3.05x10^{7}$	$2.20x10^{7}$	$1.40x10^{7}$	$1.00x10^{7}$	$1.14x10^{7}$	$9.30x10^{6}$	$8.25x10^{6}$	$7.40x10^{6}$	$1.11x10^{7}$	$7.90x10^{6}$	$6.75x10^{6}$	$6.10x10^{6}$
YGR220c	$1.29x10^{8}$	$8.80x10^{7}$	$6.90x10^{7}$	$5.75x10^{7}$	$2.10x10^{7}$	$2.70x10^{7}$	$1.90x10^{7}$	$1.11x10^{7}$	$1.04x10^{7}$	$8.15x10^{6}$	$8.65x10^{6}$	$8.10x10^{6}$	$8.30x10^{6}$
YJR113c	$7.35x10^{7}$	$1.33x10^{8}$	$9.70x10^{7}$	$4.43x10^{7}$	no data	$3.25x10^{7}$	$2.30x10^{7}$	$1.30x10^{7}$	$8.95x10^{6}$	$6.80x10^{6}$	$5.85x10^{6}$	$1.35x10^{6}$	$5.75x10^{6}$
YKR085c	$1.10x10^{8}$	$8.45x10^{7}$	$8.30x10^{7}$	$3.81x10^{7}$	$2.95x10^{7}$	$3.85x10^{7}$	$1.75x10^{7}$	$9.00x10^{6}$	$9.70x10^{6}$	$7.60x10^{6}$	$7.35x10^{6}$	$8.85x10^{6}$	$7.55x10^{6}$
YNL177c	$1.07x10^{8}$	$6.85x10^{7}$	$3.75x10^{7}$	$3.10x10^{7}$	$4.40x10^{7}$	$9.00x10^{6}$	$1.09x10^{7}$	$6.55x10^{6}$	$4.60x10^{6}$	$1.24x10^{7}$	$5.70x10^{6}$	$8.30x10^{6}$	$1.28x10^{7}$

Heterozygous diploid, washed cultures

Week	1	2	3	4	5	6	7	8	9	10	11	12	13
YCR003w	$1.68x10^{8}$	$1.10x10^{7}$	$1.65x10^{7}$	$1.20x10^{7}$	$9.50x10^{6}$	$4.60x10^{6}$	no data	no data	$1.10x10^{6}$	$1.40x10^{6}$	no data	no data	no data
YDL202w	$6.70x10^{7}$	$2.05x10^{7}$	no data	$6.50x10^{6}$	$3.00x10^{6}$	$2.40x10^{6}$	$2.30x10^{6}$	$8.00x10^{5}$	no data	$6.85x10^{6}$	$3.10x10^{5}$	$2.70x10^{5}$	$8.45x10^{6}$
YDL405w	$4.42x10^{7}$	$3.15x10^{7}$	$2.90x10^{7}$	$1.65x10^{7}$	$1.35x10^{7}$	$3.80x10^{6}$	$3.50x10^{6}$	$3.05x10^{6}$	$1.85x10^{6}$	$1.65x10^{6}$	$3.20x10^{6}$	$1.65x10^{6}$	$7.50x10^{5}$
YDR194c	$9.40x10^{2}$	$8.15x10^{2}$	$5.40x10^{2}$	$6.35x10^{2}$	$3.75x10^{2}$	$2.60x10^{2}$	$2.80x10^{2}$	$8.00x10^{1}$	$3.50x10^{1}$	$3.21x10^{3}$	$3.00x10^{2}$	$4.35x10^{2}$	$3.35x10^{2}$
YGL135c	$4.15x10^{7}$	$5.00x10^{7}$	$2.60x10^{7}$	$1.05x10^{7}$	$1.50x10^{6}$	$2.65x10^{6}$	$3.40x10^{6}$	$3.25x10^{6}$	$3.85x10^{6}$	$2.20x10^{6}$	$2.40x10^{6}$	$1.65x10^{6}$	$1.05x10^{6}$
YGR220c	$1.05x10^{8}$	$8.65x10^{7}$	$5.25x10^{7}$	$5.30x10^{7}$	$2.30x10^{7}$	$8.00x10^{6}$	$8.00x10^{6}$	$4.35x10^{6}$	$4.15x10^{6}$	$4.05x10^{6}$	$6.50x10^{6}$	$4.30x10^{6}$	$3.05x10^{6}$
YJR113c	$1.14x10^{8}$	$7.60x10^{7}$	$9.45x10^{7}$	$5.20x10^{7}$	$3.45x10^{7}$	$2.70x10^{7}$	$1.15x10^{7}$	$2.45x10^{6}$	$1.93x10^{6}$	$1.45x10^{6}$	$9.50x10^{5}$	$1.75x10^{6}$	$5.20x10^{5}$
YKR085c	$9.55x10^{7}$	$7.05x10^{7}$	$1.00x10^{8}$	$4.40x10^{7}$	$3.00x10^{7}$	$5.00x10^{6}$	$9.50x10^{5}$	$9.50x10^{5}$	$1.93x10^{6}$	$1.43x10^{6}$	$8.50x10^{5}$	$2.06x10^{6}$	$1.71x10^{6}$
YNL177c	$9.50x10^{7}$	$1.08x10^{8}$	$7.25x10^{7}$	$4.05x10^{7}$	$2.25x10^{7}$	$1.00x10^{6}$	$1.70x10^{6}$	$4.28x10^{6}$	$9.50x10^{5}$	$4.50x10^{6}$	$2.90x10^{6}$	$1.25x10^{6}$	$2.85x10^{6}$

Homozygous diploid, unwashed cultures

Week	1	2	3	4	5	6	7	8	9	10	11	12	13
YCR003w	$2.96x10^7$	$3.40x10^7$	no data	$1.65x10^7$	$1.40x10^7$	$1.96x10^7$	$1.71x10^7$	$1.57x10^7$	$1.61x10^7$	$1.22x10^7$	$7.50x10^6$	$9.85x10^6$	$8.40x10^6$
YDL202w	$7.00x10^7$	$3.60x10^7$	no data	no data	no data	$6.00x10^5$	$5.50x10^4$	$6.00x10^4$	$1.20x10^4$	$1.00x10^3$	no data	$7.50x10^1$	no data
YDL405w	$5.70x10^7$	no data	$2.50x10^6$	$3.00x10^6$	$2.00x10^6$	$4.75x10^5$	no data	$5.00x10^3$	no data	no data	no data	no data	no data
YDR194c	no data	no data	no data	no data	no data	no data	no data	no data	no data	no data	no data	no data	no data
YGL135c	$5.25x10^7$	$4.55x10^7$	$3.15x10^7$	$2.40x10^7$	$3.00x10^6$	$1.63x10^7$	$1.34x10^7$	$1.08x10^7$	$7.80x10^6$	$8.80x10^6$	$6.95x10^6$	$8.50x10^6$	$6.70x10^6$
YGR220c	$1.70x10^8$	$1.56x10^8$	$1.51x10^8$	$1.54x10^8$	$9.20x10^7$	$7.45x10^7$	$6.60x10^7$	$2.70x10^7$	$2.15x10^7$	$1.34x10^7$	$5.85x10^6$	$6.85x10^6$	no data
YJR113c	$1.55x10^8$	$8.95x10^7$	$1.65x10^7$	$1.33x10^7$	$1.75x10^7$	$8.90x10^6$	$2.45x10^6$	no data	$1.00x10^5$	$4.10x10^4$	$8.00x10^3$	$1.20x10^3$	no data
YKR085c	$1.05x10^8$	$4.10x10^7$	$1.70x10^7$	$1.50x10^7$	$1.40x10^7$	$4.25x10^6$	$5.50x10^5$	$1.95x10^5$	$1.00x10^5$	$1.50x10^4$	no data	no data	$2.00x10^1$
YNL177c	$1.60x10^8$	$1.34x10^8$	no data	$1.49x10^8$	$2.40x10^8$	no data	$2.50x10^7$	$1.70x10^7$	$6.00x10^6$	$3.55x10^6$	$4.30x10^6$	$2.30x10^6$	$1.15x10^6$

Homozygous diploid, washed cultures

Week	1	2	3	4	5	6	7	8	9	10	11	12	13
YCR003w	$3.82x10^7$	$2.85x10^7$	$2.60x10^7$	$4.00x10^6$	$4.00x10^6$	$4.55x10^6$	$2.90x10^6$	$3.10x10^6$	$1.60x10^6$	$1.65x10^6$	$6.00x10^5$	$5.55x10^6$	$2.35x10^5$
YDL202w	$8.25x10^7$	$1.20x10^7$	no data	$5.00x10^5$	$5.00x10^5$	no data	no data	no data	no data	no data	no data	no data	no data
YDL405w	no data	$4.00x10^6$	no data	$3.00x10^6$	$5.00x10^5$	no data	no data	$1.00x10^4$	no data	no data	no data	$1.00x10^1$	no data
YDR194c	no data	no data	no data	no data	no data	no data	no data	no data	no data	no data	no data	no data	no data
YGL135c	$4.95x10^7$	$3.20x10^7$	$2.80x10^7$	$1.00x10^7$	$4.00x10^6$	$3.20x10^6$	$2.25x10^6$	$3.15x10^6$	$3.55x10^6$	$1.45x10^6$	$2.15x10^6$	$1.65x10^6$	$1.85x10^6$
YGR220c	$1.69x10^8$	$1.25x10^8$	$1.35x10^8$	$1.49x10^8$	$9.90x10^7$	$7.20x10^7$	$3.75x10^7$	$1.55x10^7$	$8.50x10^6$	$5.35x10^6$	$5.85x10^6$	$1.80x10^6$	$1.45x10^6$
YJR113c	$1.30x10^8$	$2.30x10^7$	$3.50x10^6$	$5.75x10^5$	$6.50x10^4$	no data	no data	$4.00x10^1$	$4.00x10^5$	no data	no data	no data	no data
YKR085c	$1.12x10^8$	$1.25x10^7$	$1.00x10^6$	$1.15x10^6$	$2.50x10^4$	$8.72x10^5$	$7.00x10^5$	$3.25x10^7$	$4.00x10^5$	no data	$6.30x10^4$	$6.45x10^4$	$1.46x10^5$
YNL177c	$1.40x10^8$	$1.38x10^8$	$1.18x10^8$	$1.52x10^8$	$2.48x10^8$	$9.10x10^7$	$5.60x10^7$	$3.60x10^7$	$1.95x10^7$	$8.50x10^6$	$4.00x10^6$	$1.57x10^7$	$1.45x10^6$

APPENDIX 5

Allowing G-U base pairing in BLAST searches

To allow G-U base pairing to be identified by BLAST the program was modified to allow the rRNA gene G nucleotide to be matched with C and A nucleotides on the database sequence while the rRNA nucleotide U was matched with A and C nucleotides on the mRNA. This is shown in the example below, using the random sequence of ATGCTC.

Genomic sequence in database (e.g. EMBL): 5′- ATGCTC -3′

This would be from duplex DNA:
5′- ATGCTG -3′
3′- TACGAC -5′

This gives the two sequences:

DNA sequence (template stand): 5′- ATGCTG -3′

Reverse complement DNA sequence (non-template strand): 5′- CAGCAT -3′

Thus the mRNA produced from this piece of DNA would be: 5′- AUGCUG -3′

If base pairing is allowed such that A pairs with U; U pairs with A or G; G pairs with C or U; C pairs with G the mRNA would therefore be able to base pair to rRNA, where all the bases are involved in base pairing in the case (where the bases are highlighted in bold normal A-T(U) and G-C base pairing will give 100% base pairing. In the unhighlighted area base pairing will only occur if G-U base pairing is allowed):

```
mRNA 5′- AUGCUG -3′
          ||||||
rRNA 3′- UACGGU -5′
```

Comparing the rRNA sequence against the DNA and reverse complement mRNA sequences (looking for identical base similarity between the sequences):

```
rRNA      5'- UGGCAU -3'        5'- UGGCAU -3'
                ||                    ||||
mRNA      5'- AUGCUG -3'        5'- CAGCAU -3'

     (DNA sequence)     (reverse complement DNA sequence)
```

Working back form the RNA sequences to the DNA sequences deposited in databases (e.g. EMBL):

RNA	5'-AUGC-3'	
DNA template strand	3'-TACG-5'	
DNA non-template strand	5'-ATGC-3'	(sequence deposited in databases)

It can be seen that the rRNA/mRNA sequence is identical to the DNA sequences deposited in the databases (with the exception of the T to U change between DNA & RNA). Therefore comparing the rRNA sequence to the mRNA is equivalent to comparing the rRNA gene sequence to the DNA gene sequence. It can be seen therefore that a comparison of the DNA sequence and the rRNA gene sequence using BLAST will find no obvious sequence matches. Therefore these sequences cannot be used in the BLAST program to identify potential base pairing sites. Comparison of the reverse complement DNA sequence and the rRNA gene sequence does however show sequence homology. Which can be seen over the highlighted area where Watson-Crick base pairing is occurring, thus the BLAST program would be able to identify these areas. However this pattern breaks down where G-U base pairing is occurring (non-highlighted areas). To allow G-U base pairing to be identified by BLAST the program must be modified to allow

the rRNA gene G nucleotide to be matched with C and A, and the rRNA nucleotide T (U) to be matched with A and C.

Appendix 6

Computer Program

Getblastresults.pl

- Written by Elizabeth Williams, The University of Bath, 2002
- Used for extraction and tabulation of specific values, for each gene, from a BLAST result output file.

```
#!/usr/bin/perl -w

print "what file do you want to search?\n";
$file = <STDIN>;

open (FILE, $file) || die "can't open $file\n";

while (<FILE>) {

        if (/Query=\s+(.+?)\n/) {
                $query = $1;
                push @queries, $query;
        }

        if (/>(.+)/) {
                $index = 0;
                $result = $1;
```

```
        push @{ $bighash{$query}{"results"} }, $result;
}

if (/Length =\s+(\w+)/) {
        $bighash{$query}{$result}{"length"} = $1;
}

if (/Score =\s+([0-9]+)/) {
        $index++;
        $bighash{$query}{$result}{"score"}[$index] = $1;
}

if (/Expect = (([0-9]+.[0-9]+e.[0-9]+))/) {
        $bighash{$query}{$result}{"expect"}[$index] = $1;
}

if (/P = ([0-9].[0-9]+)/) {
        $bighash{$query}{$result}{"probability"}[$index] = $1;
}

if (/Identities = ([0-9]+)\/([0-9]+)/) {
        $bighash{$query}{$result}{"identities"}[$index] = $1;
        $bighash{$query}{$result}{"denominator"}[$index] = $2;
}

if (/Positives = ([0-9]+)/) {
```

```
            $bighash{$query}{$result}{"positives"}[$index] = $1;
        }

        if (/Strand = (Minus|Plus)/) {
            $bighash{$query}{$result}{"strand"}[$index] = $1;
        }

        if (/Query:\s+([0-9]+)/) {
            unless (exists $bighash{$query}{$result}{"querystart"}[$index]) {
            $bighash{$query}{$result}{"querystart"}[$index] = $1;
            }
        }
        if (/Sbjct:\s+([0-9]+)/) {
            unless (exists $bighash{$query}{$result}{"subjctstart"}[$index]) {
            $bighash{$query}{$result}{"subjctstart"}[$index] = $1;
            }
        }
}

$lenOfFileName = length($file);
$lenOfFileName = $lenOfFileName - 4;
$fileName = substr($file, 0, $lenOfFileName);
$fileNameXLS = $fileName . "xls";

#open (BEST, ">$fileNameXLS") || die "can't open $fileNameXLS\n";
```

```
#print BEST
"Query\tResults\tLength\tScore\tExpect\tProbability\tIdentity\tPositives\tLength of
Hit\tStrand\n";

open (OUT, ">$fileNameXLS") || die "can't open $fileNameXLS\n";

print OUT
"Query\tResults\tLength\tScore\tExpect\tProbability\tIdentity\tPositives\tLength of
Hit\tStrand\tquerystart\tsubjctstart\n";

foreach $queryItem (@queries) {

        #foreach $item (@{ $bighash{$queryItem}{"results"} }) {
#               $len = $bighash{$queryItem}{$item}{"length"};
#               $scr = $bighash{$queryItem}{$item}{"score"}[1];
#               $exp = $bighash{$queryItem}{$item}{"expect"}[1];
#               $prob = $bighash{$queryItem}{$item}{"probability"}[1];
#               $id = $bighash{$queryItem}{$item} {"identities"}[1];
#               $pos = $bighash{$queryItem}{$item}{"positives"}[1];
#               $den = $bighash{$queryItem}{$item}{"denominator"}[1];
#               $str = $bighash{$queryItem}{$item}{"strand"}[1];
#
                #print BEST
"$queryItem\t$item\t$len\t$scr\t$exp\t$prob\t$id\t$pos\t$den\t$str\n";
#       }

        foreach $item (@{ $bighash{$queryItem}{"results"} }) {
```

```
            for $i (1 .. $#{$bighash{$queryItem}{$item}{"score"}} ) {
                if (defined $bighash{$queryItem}{$item}{"score"}[$i]){
                    $len = $bighash{$queryItem}{$item}{"length"};
                    $scr = $bighash{$queryItem}{$item}{"score"}[$i];
                    $exp = $bighash{$queryItem}{$item}{"expect"}[$i];
                    $prob = $bighash{$queryItem}{$item}{"probability"}[$i];
                    $id = $bighash{$queryItem}{$item} {"identities"}[$i];
                    $pos = $bighash{$queryItem}{$item}{"positives"}[$i];
                    $den = $bighash{$queryItem}{$item}{"denominator"}[$i];
                    $str = $bighash{$queryItem}{$item}{"strand"}[$i];
                    $ques= $bighash{$queryItem}{$item}{"querystart"}[$i];
                    $subs= $bighash{$queryItem}{$item}{"subjctstart"}[$i];

                    print OUT
"$queryItem\t$item\t$len\t$scr\t$exp\t$prob\t$id\t$pos\t$den\t$str\t$ques\t$subs\n";
                }
            }
        }
}
```